

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXY 10
DB 201 VVPAGGSY 208

RESULT 13

US-08-206-188B-28
Sequence 28, Application US/08206188B
Patent No. 6100025
GENERAL INFORMATION:
APPLICANT: Migler, Michael H.
COLICEILL, John J.
TITLE OF INVENTION: Cloning by Complementation and Related
Processes
NUMBER OF SEQUENCES: 84
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Garstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/206,188B
FILING DATE: 01-MAR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36107
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 688 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULAR TYPE: protein
US-08-206-188B-28

Query Match 64.6%; Score 31; DB 3; Length 688;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXY 10
DB 201 VVPAGGSY 208

RESULT 14

US-08-844-086-2
Sequence 2, Application US/08844086
Patent No. 5866390
GENERAL INFORMATION:
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 5866390e1 Compounds
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA

COUNTRY: USA
ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/844,086

FILING DATE: 18-APR-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9607993.4

FILING DATE: 18-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Gimmil, Edward R

REGISTRATION NUMBER: 38,891

REFERENCE/DOCKET NUMBER: P31457-4

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-4478

TELEFAX: 610-270-5090

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 833 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULAR TYPE: protein

US-08-844-086-2

Query Match 64.6%; Score 31; DB 2; Length 833;
Best Local Similarity 66.7%; Pred. No. 3.3e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGXS 9
DB 169 BEVLPDGT 177

RESULT 15

US-09-018-211-2
Sequence 2, Application US/09018211
Patent No. 6048716
GENERAL INFORMATION:
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 6048716e1 Compounds
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,211
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/844,086
FILING DATE: 18-APR-1997
APPLICATION NUMBER: 9607993.4
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmil, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31457-4

TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 833 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-018-211-2

Query Match 64.6%; Score 31; DB 3; Length 833;
Best Local Similarity 66.7%; Pred. No. 3.3e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EEVVPXGS 9
|||:|
Db 169 EEVLDPGTS 177

Search completed: December 22, 2003, 16:43:45
Job time : 11.133 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 16:40:14 ; Search time 20.8667 Seconds
(without alignments)
98.451 Million cell updates/sec

Title: US-09-909-164-43
Perfect score: 48
Sequence: 1 BEVVPXGXSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	77.1	153	12	US-10-029-386-32486	Sequence 32486, A
2	75.0	1163	12	US-10-122-067-4	Sequence 4, Appl1
3	75.0	1499	12	US-10-096-534-67	Sequence 67, Appl1
4	75.0	1499	12	US-10-123-067-2	Sequence 2, Appl1
5	75.0	3472	12	US-10-029-120-4	Sequence 4, Appl1
6	75.0	3472	14	US-10-027-806-4	Sequence 4, Appl1
7	75.0	3472	14	US-10-034-623-4	Sequence 4, Appl1
8	75.0	3472	15	US-10-027-801-4	Sequence 4, Appl1
9	70.8	478	11	US-09-992-600A-108	Sequence 108, App
10	70.8	478	11	US-09-924-340-108	Sequence 108, App
11	70.8	478	11	US-09-746-783-184	Sequence 184, App
12	70.8	478	12	US-09-992-095B-108	Sequence 108, App
13	70.8	478	12	US-10-154-678-108	Sequence 108, App
14	70.8	478	12	US-09-999-570-108	Sequence 108, App
15	70.8	478	15	US-10-000-489-108	Sequence 108, App

16	34	70.8	478	15	US-10-000-986-108	Sequence 108, App
17	32	66.7	236	12	US-10-029-386-32076	Sequence 32076, A
18	32	66.7	426	15	US-10-094-749-2565	Sequence 2565, App
19	32	66.7	290	12	US-10-214-766-43	Sequence 43, Appl1
20	32	66.7	947	15	US-10-101-464A-73	Sequence 73, Appl1
21	31	64.6	99	15	US-10-106-698-7477	Sequence 7477, App
22	31	64.6	597	9	US-09-815-242-13226	Sequence 13226, A
23	31	64.6	653	11	US-09-820-843A-26	Sequence 26, Appl1
24	31	64.6	1083	11	US-09-835-976B-2	Sequence 2, Appl1
25	31	64.6	1083	11	US-09-835-976B-14	Sequence 14, Appl1
26	31	64.6	1099	11	US-09-835-976B-10	Sequence 10, Appl1
27	31	64.6	1116	11	US-09-835-976B-12	Sequence 12, Appl1
28	31	64.6	1135	11	US-09-835-976B-4	Sequence 4, Appl1
29	31	64.6	1135	11	US-09-835-976B-6	Sequence 6, Appl1
30	31	64.6	1150	11	US-09-835-976B-8	Sequence 8, Appl1
31	31	64.6	1150	11	US-09-835-976B-16	Sequence 16, Appl1
32	31	64.6	1152	16	US-10-080-170-130	Sequence 130, App
33	30	62.5	7	11	US-09-909-062-1	Sequence 1, Appl1
34	30	62.5	7	11	US-09-909-062-9	Sequence 9, Appl1
35	30	62.5	7	11	US-09-909-062-130	Sequence 130, App
36	30	62.5	121	9	US-09-853-161-68	Sequence 85, Appl1
37	30	62.5	121	9	US-09-853-161-85	Sequence 85, Appl1
38	30	62.5	121	9	US-09-852-659A-68	Sequence 68, Appl1
39	30	62.5	121	9	US-09-852-659A-85	Sequence 85, Appl1
40	30	62.5	121	10	US-09-852-797-85	Sequence 85, Appl1
41	30	62.5	121	10	US-09-852-797-85	Sequence 85, Appl1
42	30	62.5	135	9	US-09-989-723-359	Sequence 359, App
43	30	62.5	135	9	US-09-989-723-359	Sequence 359, App
44	30	62.5	135	9	US-09-989-723-359	Sequence 359, App
45	30	62.5	135	9	US-09-989-723-359	Sequence 359, App

ALIGNMENTS

RESULT 1
US-10-029-386-32486
; Sequence 32486, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AECOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32486
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF002994.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: SWISSPROT HIT: P43363, EVALUATE 5.00e-46
US-10-029-386-32486

Query Match 77.1%; Score 37; DB 12; Length 153;
Best Local Similarity 70.0%; Pred. No. 4.4;
Matches 7; Conservatively 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 BEVVPXGXSY 10
DB 15 KEVVPFGHSY 24

RESULT 2
US-10-122-067-4
; Sequence 4, Application US/10122067
; Publication No. US20030165883A1
; GENERAL INFORMATION:
; APPLICANT: ROY A.J. CURTIS
; TITLE OF INVENTION: 27091, A PHOSPHOLIPID TRANSPORTING
; TITLE OF INVENTION: ATPASE MOLECULE AND USES THEREFOR
; FILE REFERENCE: MP101-058PIR
; CURRENT APPLICATION NUMBER: US/10/122,067
; CURRENT FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/283434
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1163
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-122-067-4

Query Match 75.0%; Score 36; DB 12; Length 1163;
Best Local Similarity 72.7%; Pred. No. 69;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXSYS 11
DB 133 BEVVPXGXSYS 143

RESULT 3
US-10-096-534-67
; Sequence 67, Application US/10096534
; Publication No. US20030166887A1
; GENERAL INFORMATION:
; APPLICANT: The Brigham and Women's Hospital, Inc.
; APPLICANT: Yates, Karen
; APPLICANT: Mizuno, Shuichi
; APPLICANT: Glowacki, Julie
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF SKELETAL DEGENERATION CONDITIONS
; FILE REFERENCE: B0801/7244/KA/ERP
; CURRENT APPLICATION NUMBER: US/10/096,534
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: US 60/274,980
; PRIOR FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 67
; LENGTH: 1499
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-096-534-67

Query Match 75.0%; Score 36; DB 12; Length 1499;
Best Local Similarity 72.7%; Pred. No. 92;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXSYS 11
DB 469 BEVVPXGXSYS 479

RESULT 4
US-10-122-067-2
; Sequence 2, Application US/10122067
; Publication No. US20030165883A1
; GENERAL INFORMATION:
; APPLICANT: ROY A.J. CURTIS
; TITLE OF INVENTION: 27091, A PHOSPHOLIPID TRANSPORTING
; TITLE OF INVENTION: ATPASE MOLECULE AND USES THEREFOR
; FILE REFERENCE: MP101-058PIR
; CURRENT APPLICATION NUMBER: US/10/122,067
; CURRENT FILING DATE: 2002-04-12

; PRIOR APPLICATION NUMBER: 60/283434
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1499
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-122-067-2

Query Match 75.0%; Score 36; DB 12; Length 1499;
Best Local Similarity 72.7%; Pred. No. 92;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXSYS 11
DB 469 BEVVPXGXSYS 479

RESULT 5
US-10-029-120-4
; Sequence 4, Application US/10029120
; Publication No. US20030175708A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM
; FILE REFERENCE: DCCP.002A
; CURRENT APPLICATION NUMBER: US/10/029,120
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-029-120-4

Query Match 75.0%; Score 36; DB 12; Length 3472;
Best Local Similarity 54.5%; Pred. No. 2,36+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGXSYS 11
DB 2294 BEVVPXGXSYS 2304

RESULT 6
US-10-027-806-4
; Sequence 4, Application US/10027806
; Publication No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM
; FILE REFERENCE: DCCP.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-806-4

Query Match 75.0%; Score 36; DB 14; Length 3472;

Best Local Similarity 54.5%; Pred. No. 2.3e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXSYS 11
||:|:|:|:|:
Db 2294 EDVIPRGISFS 2304

RESULT 7

US-10-034-623-4
; Sequence 4, Application US/10034623
; Publication No. US20020198365A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOF. 002A
; CURRENT APPLICATION NUMBER: US/10/034,623
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-034-623-4

Query Match 75.0%; Score 36; DB 14; Length 3472;
Best Local Similarity 54.5%; Pred. No. 2.3e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXSYS 11
||:|:|:|:|:
Db 2294 EDVIPRGISFS 2304

RESULT 8

US-10-027-801-4
; Sequence 4, Application US/10027801
; Publication No. US20030054364A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOF. 002A
; CURRENT APPLICATION NUMBER: US/10/027,801
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-801-4

Query Match 75.0%; Score 36; DB 15; Length 3472;
Best Local Similarity 54.5%; Pred. No. 2.3e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXSYS 11
||:|:|:|:|:
Db 2294 EDVIPRGISFS 2304

RESULT 9

US-09-992-600A-108
; Sequence 108, Application US/0992600A
; Publication No. US20030027161A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephanie
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US4.DIV
; CURRENT APPLICATION NUMBER: US/09/992,600A
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: JPatent
; SEQ ID NO 108
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-992-600A-108

Query Match 70.8%; Score 34; DB 11; Length 478;
Best Local Similarity 60.0%; Pred. No. 69;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EEVVPXGXSYS 11
||:|:|:|:|:
Db 239 EVAPAGASYN 248

RESULT 10
US-09-924-340-108
; Sequence 108, Application US/09924340
; Publication No. US20030027248A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephanie
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US2.REG
; CURRENT APPLICATION NUMBER: US/09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 108
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-924-340-108

Query Match 70.8%; Score 34; DB 11; Length 478;
Best Local Similarity 60.0%; Pred. No. 69;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EEVVPXGXSYS 11
||:|:|:|:|:
Db 239 EVAPAGASYN 248

RESULT 11
US-09-746-783-184
; Sequence 184, Application US/09746783
; Publication No. US20030044935A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John M.
; Lavallee, Edward R.
; Racie, Lisa A.
; Treacy, Maurice
; Spaulding, Vikki
; Agostino, Michael J.
; Howes, Steven H.
; Fechtel, Kim
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; ENCODING THEM
; NUMBER OF SEQUENCES: 231
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/746,783
; FILING DATE: 21-Dec-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Milasincic, Debra J.
; REGISTRATION NUMBER: 46,931
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 184:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 478 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 184:
US-09-746-783-184
Query Match 70.8%; Score 34; DB 11; Length 478;
Best Local Similarity 60.0%; Pred. No. 69;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
CY 2 EVVPGXGXSYS 11
DB 239 EVAPAGASYN 248
RESULT 12
US-09-992-095B-108
; Sequence 108, Application US/09992095B
; Publication No. US20030157485A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephanie
; Applicant: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.053.DIV
; CURRENT APPLICATION NUMBER: US/09/992,095B
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06

; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 108
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-992-095B-108
Query Match 70.8%; Score 34; DB 12; Length 478;
Best Local Similarity 60.0%; Pred. No. 69;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
CY 2 EVVPGXGXSYS 11
DB 239 EVAPAGASYN 248
RESULT 13
US-10-154-678-108
; Sequence 108, Application US/10154678
; Publication No. US20030162166A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephanie
; Applicant: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 182. US1.REG
; CURRENT APPLICATION NUMBER: US/10/154,678
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 108
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-154-678-108
Query Match 70.8%; Score 34; DB 12; Length 478;
Best Local Similarity 60.0%; Pred. No. 69;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
CY 2 EVVPGXGXSYS 11
DB 239 EVAPAGASYN 248
RESULT 14
US-09-999-570-108
; Sequence 108, Application US/09999570
; Publication No. US20030170628A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephanie
; Applicant: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF

Job time : 20.9333 secs

```
; FILE REFERENCE: G-091US08DIV
; CURRENT APPLICATION NUMBER: US/09/999,570
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 108
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-999-570-108
```

```
Query Match 70.8%; Score 34; DB 12; Length 478;
Best Local Similarity 60.0%; Pred. No. 69;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
Oy 2 EVVPXGXSYS 11
Db 239 EVAPAGASYN 248
```

```
RESULT 15
US-10-000-489-108
; Sequence 108; Application US/10000489
; Publication No. US20030092011A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US6.DIV
; CURRENT APPLICATION NUMBER: US/10/000,489
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 108
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-000-489-108
```

```
Query Match 70.8%; Score 34; DB 15; Length 478;
Best Local Similarity 60.0%; Pred. No. 69;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
Oy 2 EVVPXGXSYS 11
Db 239 EVAPAGASYN 248
```

Search completed: December 22, 2003, 17:32:44

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 17:24:36 / Search time 9.06667 Seconds

(without alignments)
116.675 Million cell updates/sec

Title: US-09-909-164-43

Perfect score: 48

Sequence: 1 EKVVPKXGXSYS 11

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

PIR 76:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	75.0	3472	2	T31308	hypothetical 367K
2	72.9	225	2	S57810	hypothetical prote
3	70.8	102	2	A42452	VI protein - cobac
4	70.8	165	2	D69493	hypothetical prote
5	70.8	259	2	T34536	hypothetical prote
6	70.8	1028	2	AF3286	ATP-dependent DNA
7	68.8	124	1	VRLJ51	trans-regulatory s
8	68.8	427	2	E64064	colB protein - Hae
9	66.7	227	2	E75619	hypothetical prote
10	66.7	425	2	T24111	3-phosphoshikimate
11	66.7	426	2	D82163	hypothetical prote
12	66.7	670	2	S22293	zinc finger protei
13	66.7	890	2	A30481	bacteriocin BCN5 -
14	66.7	2717	2	A34203	DNA-binding protei
15	64.6	123	3	B69342	conserved hypotet
16	64.6	284	2	S75817	hypothetical prote
17	64.6	319	2	S03833	hypothetical prote
18	64.6	361	2	S15299	dtDP-glucose 4,6-de
19	64.6	361	2	AF0767	hypothetical prote
20	64.6	437	2	AG2945	periplasmic sorbit
21	64.6	450	2	C98337	ATP-dependent DNA
22	64.6	541	2	AH2679	probable ABC subst
23	64.6	544	2	C82900	type II secretion
24	64.6	561	2	C84239	DNA ligase (AB0425
25	64.6	573	2	F97461	hypothetical prote
26	64.6	612	2	T05331	hypothetical prote
27	64.6	622	2	S15009	probable beta-gala
28	64.6	646	2	C95978	iron(III) ABC tran
29	64.6	653	2	D82352	

30	31	64.6	701	2	S61239	hypothetical prote
31	31	64.6	829	2	H86726	leucine-tRNA ligas
32	31	64.6	833	2	H95029	leucyl-tRNA synthe
33	31	64.6	833	2	C97901	leucine-tRNA ligas
34	31	64.6	840	2	T39116	probable sulfate p
35	31	64.6	877	2	T40413	sulfate permease -
36	31	64.6	926	3	AG1860	hypothetical prote
37	31	64.6	1014	2	T17275	hypothetical prote
38	31	64.6	1081	1	A42399	hololeucine-tRNA li
39	31	64.6	1085	2	T18369	K-CI cotransport p
40	31	64.6	1085	2	T31429	K-CI cotransport p
41	31	64.6	1086	2	T14114	K-CI cotransport p
42	31	64.6	1088	2	D82246	probable chitinase
43	31	64.6	1116	2	T31432	K-CI cotransport p
44	31	64.6	1152	2	D87046	conserved hypotet
45	31	64.6	1548	2	T04456	hypothetical prote

ALIGNMENTS

RESULT 1
T31308
hypothetical 367K protein - Cenarchaeum symbiosum
C/Species: Cenarchaeum symbiosum
C/Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C/Accession: T31308
R/Schleper, C.; DeLong, E.F.; Preston, C.M.; Feldman, R.A.; Wu, K.Y.; Swanson, R.V.
J. Bacteriol. 180, 5003-5009, 1998
A/Title: Genomic analysis reveals chromosomal variation in natural populations of the
A/Reference number: Z20994; MUID:98422450; PMID:9748430
A/Accession: T31308
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-3472 <SCH>
A/Cross-references: EMBL:AF083072; NID:9359393; PID:9359394; PIDN:ACC6269.1
C/Superfamily: Cenarchaeum symbiosum hypothetical 367K protein

Query Match 75.0%; Score 36; DB 2; Length 3472;
Best Local Similarity 54.5%; Pred. NO. 79;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKVVPKXGXSYS 11
DB 2294 EDVPRGISPS 2304

RESULT 2
S57810
hypothetical protein precursor (clone TPPI1) - tomato
C/Species: Lycopersicon esculentum (tomato)
C/Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
C/Accession: S57810
R/Mulligan, S.B.; Gasser, C.S.
Plant Mol. Biol. 28, 691-711, 1995
A/Title: Nature and regulation of plant-expressed genes in tomato.
A/Reference number: S57808; MUID:95375233; PMID:7647301
A/Accession: S57810
A/Status: preliminary; nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-225 <ML>
A/Cross-references: EMBL:U20592; NID:9924625; PIDN:AAA80497.1; PID:9924626
C/Superfamily: plant Kunitz-type proteinase inhibitor

Query Match 72.9%; Score 35; DB 2; Length 225;
Best Local Similarity 54.5%; Pred. NO. 7;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKVVPKXGXSYS 11
DB 32 EDVVPNGKTYA 42

RESULT 3

A42452

V1: protein - tobacco yellow dwarf virus (strain Australia)

C/Species: tobacco yellow dwarf virus

C/Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999

C/Accession: A42452

R/Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.B.

Virology 187, 633-642, 1992

A/Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow

A/Reference number: A42452; MUID:92188538; PMID:1546458

A/Accession: A42452

A/Molecule type: DNA

A/Residues: 1-102 <MOR>

A/Cross-references: GB:M81103; NID:g335283; PIDN:AAA47947.1; PID:g335284

Query Match

Best Local Similarity 70.8%; Score 34; DB 2; Length 102;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVFXGXSYS 11

DB 7 QVVPFGAGSY 16

RESULT 4

D69493

hypothetical protein AF1949 - Archaeoglobus fulgidus

C/Species: Archaeoglobus fulgidus

C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 04-Mar-2000

C/Accession: D69493

R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson

J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirschner, E.F.

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A/Authors: Utecher, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S.

Smith, H.O.; Woese, C.R.; Venter, J.C.

A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae

A/Reference number: A69250; MUID:98049343; PMID:9389475

A/Accession: D69493

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-165 <KLB>

A/Cross-references: GB:AE000968; GB:AE000782; NID:g2689291; PIDN:AB89307.1; PID:g264854

C/Superfamily: Archaeoglobus fulgidus hypothetical protein AF1949

Query Match

Best Local Similarity 70.8%; Score 34; DB 2; Length 165;

Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVFXGXSYS 10

DB 60 EESIPDGASY 69

RESULT 5

T34536

hypothetical protein DKFZp434C031.1 - human (fragment)

C/Species: Homo sapiens (man)

C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C/Accession: T34536

R/Poulik, A.; Wellenreuther, R.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, October 1999

A/Reference number: Z21540

A/Accession: T34536

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-259 <POU>

A/Cross-references: EMBL:AL122063

C/Experimental source: adult testis; clone DKFZp434C031

A/Note: DKFZp434C031.1

Query Match

70.8%; Score 34; DB 2; Length 259;

Best Local Similarity 60.0%; Pred. No. 13;

Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVFXGXSYS 11

DB 22 EVVAPGAGSY 31

RESULT 6

AF3286

ATP-dependent DNA helicase BMEI0275 [imported] - Brucella melitensis (strain 16M)

C/Species: Brucella melitensis

C/Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002

C/Accession: AF3286

R/Dalvacchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mufer, C.; Log, T.; Ivanova

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A/Title: The genome sequence of the facultative intracellular pathogen Brucella melite

A/Reference number: AD3252; PMID:11756688

A/Accession: AF3286

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1028 <KUR>

A/Cross-references: GB:AB008917; PIDN:AAL51457.1; PID:g17982167; GSPDB:GM00190

A/Experimental source: strain 16M

C/Genetics:

A/Map position: 1

Query Match

Best Local Similarity 70.8%; Score 34; DB 2; Length 1028;

Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVFXGXSYS 11

DB 76 EKVPPGAGSY 86

RESULT 7

VKLJ51

trans-regulatory splicing protein - simian immunodeficiency virus SIVcpz

N/Alternate names: anti-repression trans-activator; art protein; rev protein; trs prot

C/Species: simian immunodeficiency virus SIVcpz

A/Note: host Pan troglodytes (chimpanzee)

C/Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999

C/Accession: S09988

R/Huet, T.; Cheyrier, R.; Meyers, A.; Roelants, G.; Wain-Hobson, S.

Nature 345, 356-359, 1990

A/Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.

A/Reference number: S09983; MUID:90259077; PMID:2188136

A/Accession: S09988

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-124 <HUB>

A/Cross-references: EMBL:X52154; NID:g58866; PIDN:CAA36405.1; PID:g763085

C/Genetics:

A/Genes: rev; trs; art

A/Intons: 27/1

C/Superfamily: AIDS trans-regulatory splicing protein; transcription regulation

Query Match

Best Local Similarity 68.8%; Score 33; DB 1; Length 124;

Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVFXGXSYS 11

DB 107 ETVVAPGAGSY 116

RESULT 8

F64064

tolB protein - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: F64064; J05213
R:Freischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
R:Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.; Glodek, A.; Kelley, J.M.; Weisman, J.
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhman, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Author: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: F64064
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-427 <TIGR>
A:Cross-references: GB:U32722; GI:142023; NID:G1573348; PIDN:AAC22040.1; PID:G1573352; T
R:Experimental source: strain Rd KW20
R:Sen, K.; Sikkema, D.J.; Murphy, T.F.
Gene 178, 75-81, 1996
A:Title: Isolation and characterization of the Haemophilus influenzae coli, coli A
A:Reference number: J05212; MUID:97080550; PMID:9921895
A:Accession: J05212
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5, 'H', 'I', '15-16', 'ITR', '20', 'V', '22-78', 'H', '80-128', 'A', '130-159', 'G', '161-236'
A:Cross-references: GB:U32470; NID:G1685076; PIDN:AAC44597.1; PID:G1685080
A:Experimental source: strain 1479
C:Genetics:
A:Gene: coliB
C:Function:
A:Description: involved in transport of colicins and phages across the cell envelope; pJ

Query Match 68.8%; Score 33; DB 2; Length 427;
Best Local Similarity 60.0%; Pred. No. 38;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 EVVFXGXSY 11
Db 103 QVPSGNGYS 112

RESULT 9
E75619
hypothetical protein DRB0013 - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: E75619
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.V.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uitterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: E75619
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-227 <WHI>
A:Cross-references: GB:AE001826; NID:G6460827; PIDN:AAI12657.1; PID:G6460953; TIGR:DRB00
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRB0013
A:Map position: megaplasmid
A:Note: plasmid MPl
C:Superfamily: Deinococcus radiodurans megaplasmid hypothetical protein DRB0013

Query Match 66.7%; Score 32; DB 2; Length 227;
Best Local Similarity 54.5%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVFXGXSY 11
Db 43 BEVLPFGSFS 53

RESULT 10
T24111
hypothetical protein R10D12.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24111
R:Percy, C.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19842
A:Accession: T24111
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-425 <WLI>
A:Cross-references: EMBL:Z81109; PIDN:CAB03241.1; GSPDB:GN00023; CESP:R10D12.10
A:Experimental source: clone R10D12
C:Genetics:
A:Gene: CESP:R10D12.10
A:Map position: 5
A:introns: 23/3; 56/3; 113/3; 257/2

Query Match 66.7%; Score 32; DB 2; Length 425;
Best Local Similarity 50.0%; Pred. No. 62;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVFXGXSY 10
Db 335 RQVPGGLQY 344

RESULT 11
D82163
3-phosphoshikimate 1-carboxyvinyltransferase VC1732 [imported] - Vibrio cholerae (str
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: D82163
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
charson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers,
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: D82163
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-426 <HEI>
A:Cross-references: GB:AE004251; GB:AE003852; NID:G9656248; PIDN:AAF94882.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1732
A:Map position: 1
C:Superfamily: 3-phosphoshikimate 1-carboxyvinyltransferase; 3-phosphoshikimate 1-car

Query Match 66.7%; Score 32; DB 2; Length 426;
Best Local Similarity 60.0%; Pred. No. 62;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVFXGXSY 10
Db 223 BEVFPAGQSY 232

RESULT 12
S22293
zinc finger protein AT-BP2 - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 20-Sep-1999
C:Accession: S22293; I78656
R:Mitchellmore, C.; Traboni, C.; Cortese, R.
Nucleic Acids Res. 19, 141-147, 1991
A:Title: Isolation of two cDNAs encoding zinc finger proteins which bind to the alpha
A:Reference number: I58280; MUID:91187610; PMID:1901405
A:Accession: S22293

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-670 <MT>

A:Cross-references: EMBL:X54250; NID:G57519; PIDN:CAA38151.1; PID:G57520

A>Note: the authors did not translate the codon for residue 1

C:Superfamily: HIV-BP2 enhancer-binding protein

C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match

Best Local Similarity 66.7%; Score 32; DB 2; Length 670;
Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXSYS 11
Db 376 VVPAGLTYS 384

RESULT 13

A30481

bacteriocin BCNS - Clostridium perfringens plasmid pIP404

C:Species: Clostridium perfringens

C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 24-Nov-1999

C:Accession: A30481; S03779

R:Garnier, T.; Cole, S.T.

J. Bacteriol. 168, 1189-1196, 1986

A>Title: Characterization of a bacteriocinogenic plasmid from Clostridium perfringens at

A:Reference number: J0354; MUID:87057020; PMID:2877971

A:Accession: A30481

A:Molecule type: DNA

A:Residues: 1-890 <GAR>

A:Cross-references: GB:M32882; GB:J03309; NID:G150738; PIDN:AAA98249.1; PID:G150739

C:Genetics:

A:Gene: bcn

A:Genome: plasmid

C:Superfamily: Clostridium perfringens plasmid pIP404 bacteriocin BCNS

C:Keywords: bacteriocin

Query Match

Best Local Similarity 66.7%; Score 32; DB 2; Length 890;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVFXGSXY 10
Db 170 EVVPGGPTY 178

RESULT 14

A34203

DNA-binding protein PRDIT-BP1 - human

N:Alternate names: major histocompatibility complex enhancer-binding protein 1

C:Species: Homo sapiens (man)

C>Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 20-Sep-1999

C:Accession: A34203; A34779

R:Pan, C.M.; Mantalaris, T.

Genes Dev. 4, 29-42, 1990

A>Title: A DNA-binding protein containing two widely separated zinc finger motifs that

A:Reference number: A34203; MUID:90169514; PMID:2106471

A:Accession: A34203

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2717 <PAN>

A:Cross-references: EMBL:X51435; NID:G38017; PIDN:CAA35798.1; PID:G38018

R:Baldwin Jr., A.S.; LeClair, K.P.; Singh, H.; Sharp, P.A.

Mol. Cell. Biol. 10, 1406-1414, 1990

A>Title: A large protein containing zinc finger domains binds to related sequence element

A:Reference number: A34779; MUID:90205817; PMID:2108316

A:Accession: A34779

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 801-1072; 'N', 1074-1168; 'K', 1170-1225; 'V', 1227-1434; 'N', 1436-1607; 'T', 1609-16

C:Superfamily: HIV-BP2 enhancer-binding protein

C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match

Best Local Similarity 66.7%; Score 32; DB 2; Length 2717;
Best Local Similarity 66.7%; Pred. No. 4.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXSYS 11
Db 2405 VVPAGLTYS 2413

RESULT 15

B69342

conserved hypothetical protein AF0741 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Dec-2002

C:Accession: B69342

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dods

; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirschner, B.

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Usterback, T.; Cotton, M.D.; Sprigge, T.; Artlich, P.; Kaine, B.P.; Sykes,

Smith, H.O.; Moese, C.R.; Venter, J.C.

A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archa

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: B69342

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-123 <KLE>

A:Cross-references: GB:AE001054; GB:AE000782; NID:G2689377; PIDN:AB90501.1; PID:G2649

C:Superfamily: uncharacterized conserved protein

Query Match

Best Local Similarity 50.0%; Score 31; DB 2; Length 123;
Best Local Similarity 50.0%; Pred. No. 27;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EEVVPXGXSXY 10
Db 51 EKIAPYGDSY 60

Search completed: December 22, 2003, 17:45:00
Job time: 10.0667 secs

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB051358; BAB47392.1; -
 DR EMBL; AY029504; AAK3100.1; -
 DR EMBL; AY029487; AAK3100.1; JOINED.
 DR EMBL; AY029488; AAK3100.1; JOINED.
 DR EMBL; AY029489; AAK3100.1; JOINED.
 DR EMBL; AY029490; AAK3100.1; JOINED.
 DR EMBL; AY029491; AAK3100.1; JOINED.
 DR EMBL; AY029492; AAK3100.1; JOINED.
 DR EMBL; AY029493; AAK3100.1; JOINED.
 DR EMBL; AY029494; AAK3100.1; JOINED.
 DR EMBL; AY029495; AAK3100.1; JOINED.
 DR EMBL; AY029496; AAK3100.1; JOINED.
 DR EMBL; AY029497; AAK3100.1; JOINED.
 DR EMBL; AY029498; AAK3100.1; JOINED.
 DR EMBL; AY029499; AAK3100.1; JOINED.
 DR EMBL; AY029500; AAK3100.1; JOINED.
 DR EMBL; AY029501; AAK3100.1; JOINED.
 DR EMBL; AY029502; AAK3100.1; JOINED.
 DR EMBL; AB011138; BAA25492.1; JOINED.
 DR GeneW; HGNC:13547; ATP10C.
 DR MIM; 605855; -
 DR MIM; 105830; -
 DR GO; GO:0016021; C:integral to membrane; NAS.
 DR GO; GO:0004012; F:phospholipid-transporting ATPase activity; NAS.
 DR GO; GO:0008360; P:regulation of cell shape; NAS.
 DR InterPro; IPR001757; ATPase_E1-E2.
 DR InterPro; IPR005639; Flippase.
 DR InterPro; IPR005834; Hydrolase.
 DR Pfam; PF00702; Hydrolase; 1.
 DR PRINTS; PR00119; CATATPASE.
 DR TIGRFAMs; TIGR01652; ATPase_Plipid; 1.
 DR TIGRFAMs; TIGR01494; ATPase_P-type; 6.
 DR PROSITE; PS00154; ATPASE_E1_E2; 1.
 KM Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
 KM Multigene family.
 FT DOMAIN 1 86
 FT TRANSMEM 87 106
 FT TRANSMEM 107 110
 FT TRANSMEM 111 128
 FT TRANSMEM 129 309
 FT TRANSMEM 310 332
 FT TRANSMEM 333 362
 FT TRANSMEM 363 384
 FT TRANSMEM 385 1087
 FT TRANSMEM 1088 1108
 FT TRANSMEM 1109 1119
 FT TRANSMEM 1120 1140
 FT TRANSMEM 1141 1170
 FT TRANSMEM 1171 1192
 FT TRANSMEM 1193 1199
 FT TRANSMEM 1200 1232
 FT TRANSMEM 1233 1228
 FT TRANSMEM 1229 1249
 FT TRANSMEM 1250 1267
 FT TRANSMEM 1268 1292
 FT TRANSMEM 1293 1499
 FT MOD_RES 427 427
 FT MOD_RES 1031 1031
 FT METAL 1035 1035
 FT METAL 467 470
 FT CONFLICT 388 388
 FT SEQUENCE 1499 AA; 167687 MW; D4996A4D0655A6D CRC64;

Query Match 75.0%; Score 36; DB 1; Length 1499;
 Best Local Similarity 72.7%; Pred. NO. 13;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 BEVVPXGXSYS 11
 DB 469 BEVVPXGXSYS 479
 RESULT 2
 CARB_FUSNN STANDARD; PRT; 1058 AA.
 AC 08RG85;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl)-
 DE phosphate synthetase ammonia chain).
 GN CARB OR F0422.
 OS Fusbacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusbacteria; Fusbacteriales; Fusbacteriaceae;
 OC Fusbacterium.
 OX NCBI_TaxID=76856;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 25586;
 RX MEDLINE=2186394; PubMed=11889109;
 RA Kapetral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 RA Bhattacharya A., Bartman A., Gardner W., Grecklin G., Zhu L.,
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman B., Bernal A.,
 RA Larsen N., D'Souza W., Malinas T., Pusch G., Haselkorn R.,
 RA Ponsstein M., Kyprides N., Overbeek R.;
 RT "Genome sequence and analysis of the oral bacterium Fusbacterium
 RT nucleatum strain ATCC 25586."
 RL J. Bacteriol. 184:2005-2018(2002).
 CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
 CC phosphate + L-glutamate + carbamoyl phosphate (By similarity).
 CC -1- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
 CC -1- PATHWAY: Arginine biosynthesis.
 CC -1- PATHWAY: Pyrimidine biosynthesis; first step.
 CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain
 CC promotes the hydrolysis of glutamine to ammonia, which is used by
 CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
 CC similarity).
 CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB010554; AAL94625.1; ALT_INIT.
 DR HAMAP; MF_01210; -; 1.
 DR InterPro; IPR006275; CARA_L_glu.
 DR InterPro; IPR005483; CPase_L.
 DR InterPro; IPR005479; CPase_L_D2.
 DR InterPro; IPR005480; CPase_L_D3.
 DR InterPro; IPR005481; CPase_L_N.
 DR InterPro; IPR004362; MGS_like.
 DR Pfam; PF00289; CPase_L_chain; 2.
 DR Pfam; PF02786; CPase_L_D2; 2.
 DR Pfam; PF02787; CPase_L_D3; 1.
 DR Pfam; PF02142; MGS; 1.
 DR PRINTS; PR00098; CPASR.
 DR TIGRFAMs; TIGR01369; CPaseL1_lrg; 1.
 DR PROSITE; PS00866; CPASR_1; 2.
 DR PROSITE; PS00867; CPASR_2; 2.
 DR Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 KM ATP-binding; Manganese; Complete proteome.
 FT DOMAIN 1 401
 FT CARBOXYPHOSPHATE SYNTHETIC DOMAIN.

```

FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT FT 547 929 CARBOXYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1058
FT NP BIND 153 210 ATP (POTENTIAL).
FT NP BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY)
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1058 AA; 117451 MW; ED7037A77C1E39F CRC64;

Query Match 72.9%; Score 35; DB 1; Length 1058;
Best Local Similarity 60.0%; Pred. No. 15;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPEXGXSYS 11
DB 190 EIVPEGLNYS 199

RESULT 3
Y1K TYDVA STANDARD; PRT; 102 AA.
AC P31619;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Hypothetical 11.2 kDa protein.
GN V1.
OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
ON NCBI_TaxID=31599;
RX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9218838; PubMed=1546458;
RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;
RT "The nucleotide sequence of the infectious cloned DNA component of
RT tobacco yellow dwarf virus reveals features of geminiviruses
RT infecting monocotyledonous plants."
RL Virology 187:633-642(1992).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its use by
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce@isb-sib.ch).
CC CC or send an email to license@isb-sib.ch.
CC -----
CC CC EMBL; M81103; AAA47947.1; -.
CC DR PIR; A42452; A42452.
CC DR InterPro; IPR002621; Gemini_mov.
CC DR Pfam; PF01708; Gemini_mov.1.
CC KW Hypothetical protein.
SQ SEQUENCE 102 AA; 11178 MW; A40BCF1E0AF55B67 CRC64;

Query Match 70.8%; Score 34; DB 1; Length 102;
Best Local Similarity 60.0%; Pred. No. 2.1;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPEXGXSYS 11
DB 7 QVPEGLNYS 16

RESULT 4
Y349_ARCFU STANDARD; PRT; 165 AA.
ID Y349_ARCFU
NC O26330;
DT 28-FEB-2003 (Rel. 41, Created)
```

```

DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein AF1949.
GN AF1949.
OS Archaeoglobus fulgidus.
OC Archaeae; Buryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
CX NCBI_TextID=2234;
KM
LN
RP [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49559;
RX MEDLINE=96049343; PubMed=9389475;
RA Klank H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Griham D.B., Kyrtides N.C.,
RA Fitzsimmons R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirschner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Goehyne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.M., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB000968; AAB89307.1; -
DR PIR; D69493; D69493.
DR TIGR; AF1949; -.
DR Hypothetical protein; Transmembrane; Complete proteome.
KW TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 141 161
SQ SEQUENCE 165 AA; 17588 MM; BEC17054810ADB8 CRC64;
Query Match 70.8%; Score 34; DB 1; Length 165;
Best Local Similarity 60.0%; Pred. No. 3.5;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 EEVYPXGXSY 10
DB 60 BESIPDGASY 69
GLTSC2.
RESULT 5
GSR2_HUMAN STANDARD; PRT; 478 AA.
ID GSR2_HUMAN
AC Q9NZM5; Q9BTC6; Q9HAX6; Q9NP1; Q9NPR4; Q9UFI2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glioma tumor suppressor candidate region gene 2 protein (p60).
GN GLTSC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
CX NCBI_TextID=9606;
[1]
RP SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.
RX MEDLINE=20175430; PubMed=10708517;
RA Smith J.S., Tachibana I., Pohl U., Lee H.K., Thararajasingam U.,
RA Portier B.P., Ueki K., Billings S., Ramaswamy S., Mohrenweiser H.W.,
RA Scheinhaber B.W., Louis D.N., Jenkins R.B.,
RA "A transactin map of the chromosome 19q-Arm glioma tumor suppressor
RT region.";
```

RL Genomics 64:44-50(2000).
 [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins P.S., Wagner L., Sherman C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
 RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueidi T.B., Toibiyuki S., Canninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhally S.J.,
 RA Bobak S.A., McEwan P.C., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulty S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Pahey J., Helton B., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schultz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallie D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [3]
 RN SEQUENCE OF 9-478 FROM N.A., AND SUBCELLULAR LOCATION.
 RP MEDLINE=99214318; PubMed=10196275;
 RA Bruni R., Fineschi B., Ogile W.O., Roizman B.;
 RT "A novel cellular protein, p60, interacting with both herpes simplex
 virus 1 regulatory proteins ICP22 and ICP0 is modified in a
 cell-type-specific manner and is recruited to the nucleus after
 infection.";
 RL J. Virol. 73:3810-3817(1999).
 [4]
 RN SEQUENCE OF 12-478 FROM N.A.
 RP Andreu N., Ercivilla X., Becarreller M., Sunmy L.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 [5]
 RP SEQUENCE OF 218-477 FROM N.A.
 RC TISSUE=Testis;
 RA Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: Interacts with HSV-1 early proteins ICP22 and ICP0.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: Expressed at high levels in heart and
 pancreas, moderate levels in placenta, liver, skeletal muscle, and
 kidney, and low levels in brain and lung.
 CC -1- SIMILARITY: BELONGS TO THE GLUTSCR2 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF182076; AAF62873.1; -
 DR EMBL; BC004229; AAH04229.1; -
 DR EMBL; BC006311; AAH06311.1; -
 DR EMBL; BC010095; AAH10095.1; -
 DR EMBL; AF296124; AAH30413.1; -
 DR EMBL; AL359335; CAB94786.1; -
 DR EMBL; AL359336; CAB94787.1; -
 DR EMBL; AL122063; CAB59242.1; -
 DR SWISS-2DPAGE; Q9NZM5; HUMAN.
 DR Genew; HGNC:4333; GLUTSCR2.
 DR MIM; 605691; -
 DR GO; GO:0008181; P:tumor suppressor; TAS.
 KW Nuclear protein; Polymorphism.
 FT VARIANT 389 389 R -> Q.
 FT /FTID=VAR_011486.

FT CONFLICT 4 6
 FT CONFLICT 9 9
 FT CONFLICT 146 191
 FT
 FT CONFLICT 198 215
 FT CONFLICT 235 235
 FT CONFLICT 417 417
 FT CONFLICT 433 477
 FT
 FT CONFLICT 434 478
 FT
 SQ SEQUENCE 478 AA; 54417 MW; 7F18923B348CB52B CRC64;
 Query Match 70.8%; Score 34; DB 1; Length 478;
 Best Local Similarity 60.0%; Pred. No. 11;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 EVVFXGXSYS 11
 DB 239 EVAPAGASYN 248

RESULT 6
 REV_SIVCZ STANDARD; PRT; 124 AA.
 ID REV_SIVCZ
 AC P17280;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE REV protein (Anti-repression transactivator protein) (ART/TRS).
 GN REV.
 OS Chimpanzee immunodeficiency virus (SIVcpz) (CIV).
 CC Viruses; Retroviruses; Retroviridae; Lentiviruses.
 OX NCBI_TaxID=11723;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90259077; PubMed=2188136;
 RA Huot T., Cheynier R., Meyerhans A., Roelants G., Wain-Hobson S.;
 RT "Genetic organization of a chimpanzee lentivirus related to HIV-1.";
 RL Nature 345:356-359(1990).
 CC -1- FUNCTION: REV APPEARS TO ACT POST-TRANSCRIPTIONALLY TO RELIEVE
 NEGATIVE REPRESSION OF GAG AND ENV PRODUCTION.
 CC -1- SUBCELLULAR LOCATION: Nuclear; accumulates in the nucleoli.
 CC -1- PTM: PHOSPHOPROTEIN WHOSE STATE OF PHOSPHORYLATION IS MEDIATED
 BY A SPECIFIC SERINE KINASE ACTIVITY PRESENT IN THE NUCLEUS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X52154; CAA36405.1; -
 DR PIR; S09988; VKLJST.
 DR HIV; X52154; REVSCP2.
 DR InterPro; IPR000625; REV_protein.
 DR Pfam; PF00424; REV_1.
 KW Transcription regulation; AIDS; Phosphorylation; Nuclear protein.
 SQ SEQUENCE 124 AA; 13701 MW; F5877D1BD5F5A7B2 CRC64;
 Query Match 68.8%; Score 33; DB 1; Length 124;
 Best Local Similarity 60.0%; Pred. No. 4.3;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 EVVFXGXSYS 11
 DB 107 ETVFAGSNYS 116

RESULT 7
TOB HAIRIN STANDARD; PRT; 427 AA.
ID TOB HAIRIN STANDARD; PRT; 427 AA.
AC P4677; P94811;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE TOB protein precursor.
DN TOB OR H10382.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_Taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KM20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirtress E.F.,
McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,
Weidman J.F., Phillips C.A., Spriggs T., Hedblom B., Cotton M.D.,
Utterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
Pine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
Graham C.L., McDonald L.A., Small K.V., Frazer C.M., Smith H.O.,
Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1479;
RX MEDLINE=97080550; PubMed=8921895;
RA Sen K., Sikkema D.J., Murphy T.F.;
RT "Isolation and characterization of the Haemophilus influenzae toL,
toR, toJ, and toB genes.";
RL Gene 178:75-81(1996).
CC -1- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
(BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE TOB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U3722; AAC22040.1; -;
DR EMBL; U3470; AAC44597.1; -;
DR PIR; F64064; F64064.
DR HSP; P19935; ICRZ.
DR TIGR; H10382; -;
DR Pfam; PF04052; TOB_N; 1.
KM Transport; Protein transport; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 427 TOB PROTEIN.
FT VARIANT 6 6 R -> H (IN STRAIN 1479).
FT VARIANT 14 14 V -> I (IN STRAIN 1479).
FT VARIANT 17 19 VGS -> ITH (IN STRAIN 1479).
FT VARIANT 21 21 R -> H (IN STRAIN 1479).
FT VARIANT 79 79 R -> V (IN STRAIN 1479).
FT VARIANT 129 129 T -> A (IN STRAIN 1479).
FT VARIANT 160 160 A -> G (IN STRAIN 1479).
FT VARIANT 237 237 A -> T (IN STRAIN 1479).
FT VARIANT 322 322 S -> N (IN STRAIN 1479).
FT VARIANT 326 326 A -> V (IN STRAIN 1479).
FT VARIANT 328 328 A -> S (IN STRAIN 1479).
SQ SEQUENCE 427 AA; 44967 MW; 0882201AEB9254B9 CRC64;

Query Match 68.8%; Score 33; DB 1; Length 427;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Qy 2 EYVVPXGXSYS 11
Db 103 QVVPXGXSYS 112

RESULT 8
AROA VIBCH STANDARD; PRT; 426 AA.
ID AROA VIBCH STANDARD; PRT; 426 AA.
AC Q9KRB0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
DN enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
GN AROA OR VC1732.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_Taxid=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heldelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
Dolson R.J., Haft D.H., Hickey B.K., Peterson J.D., Umayam L.A.,
Gall S.R., Nelson K.E., Read T.D., Tettein H., Richardson D.,
Ermoslavova M.D., Vamathevan J., Bass S., Qin H., Dracot I., Sellers P.,
McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae.";
RL Nature 406:477-483(2000).
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC sixth step.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the EPSP synthase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AR004251; AAP94882.1; -;
DR PIR; D82163; D82163.
DR TIGR; VC1732; -;
DR HAMAP; MF 00210; -; 1.
DR InterPro; IPR006264; AROA.
DR InterPro; IPR001986; EPSP synthase.
DR Pfam; PF00275; EPSP synthase; 1.
DR ProDom; PD001867; EPSP synthase; 1.
DR TIGRfam; TIGR01356; aroA; 1.
DR PROSITE; PS00104; EPSP SYNTHASE_1; 1.
DR PROSITE; PS00885; EPSP SYNTHASE_2; 1.
KM Aromatic amino acid biosynthesis; Transferrase; Complete proteome.
SQ SEQUENCE 426 AA; 46101 MW; 38852D6483F81C3 CRC64;
Qy 1 EYVVPXGXSYS 10
Query Match 66.7%; Score 32; DB 1; Length 426;
Best Local Similarity 60.0%; Pred. No. 26;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 223 BPV1PAGQSY 232

RESULT 9
BCNS_CLOPE STANDARD; PRT; 890 AA.
ID BCNS_CLOPE
AC P08636;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Bacteriocin BCNS.
BN BGN.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CPN50;
RX MEDLINE=86336297; PubMed=2901768;
RA Garner T., Cole S.T.;
RT "Complete nucleotide sequence and genetic organization of the
RT bacteriocinogenic plasmid, pIP404, from Clostridium perfringens.";
RL Plasmid 19:134-150(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CPN50;
RX MEDLINE=87057020; PubMed=2877971;
RA Garner T., Cole S.T.;
RT "Characterization of a bacteriocinogenic plasmid from Clostridium
RT perfringens and molecular genetic analysis of the
RT bacteriocin-encoding gene.";
RL J. Bacteriol. 168:1189-1196(1986).
RN [3]
RP SEQUENCE OF 1-14 FROM N.A.
RC STRAIN=CPN50;
RX MEDLINE=89039249; PubMed=2460717;
RA Garner T., Cole S.T.;
RT "Studies of UV-inducible promoters from Clostridium perfringens in
RT vivo and in vitro.";
RL Mol. Microbiol. 2:607-614(1988).
CC -1- FUNCTION: MAY FUNCTION AS AN IONOPHORE.
CC -1- INDUCTION: BY UV IRRADIATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M14481; AAA98248.1; -;
DR EMBL; M32882; AAA98249.1; -;
DR PIR; A30481; A30481.
DR InterPro; IPR003646; SH3_bac.
DR InterPro; IPR000834; Zn_Carboxypept.
DR Pfam; PF00246; Zn_Carboxypept; 1.
DR SMART; SM00287; SH3b; 3
KW Antibiotic; Bacteriocin; Plasmid.
FT DOMAIN 815 869 HYDROPHOBIC.
SQ SEQUENCE 890 AA; 96699 MW; P4S5B9971C31C6C CRC64;
Query Match 66.7%; Score 32; DB 1; Length 890;
Best Local Similarity 66.7%; Pred. NO. 56;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPGXSY 10
DB 1.70 EVVPGXSY 178

RESULT 10

ZEP1_HUMAN STANDARD; PRT; 2717 AA.

ID ZEP1_HUMAN
AC P15832;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 140 (Human immunodeficiency virus type I enhancer-
DE binding protein 1) (HIV-BP1) (Major histocompatibility complex binding
DE protein 1) (MBP-1) (Positive regulatory domain II binding factor 1)
DE (PREDI-BP1).
BN HIVBP1 OR ZNF40.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90169514; PubMed=2106471;
RA Fan G.M., Maniatis T.;
RT "A DNA-binding protein containing two widely separated zinc finger
RT motifs that recognize the same DNA sequence.";
RL Genes Dev. 4:29-42(1990).
RN [2]
RP STRUCTURE BY NMR OF 2113-2142.
RX MEDLINE=91064333; PubMed=2248949;
RA Omichinski J.G., Clore G.M., Appella E., Saksaguchi K.,
RA Gronenborn A.M.;
RT "High-resolution three-dimensional structure of a single zinc finger
RT from a human enhancer binding protein in solution.";
RL Biochemistry 29:9324-9334(1990).
RN [3]
RP STRUCTURE BY NMR OF 2087-2142.
RX MEDLINE=92232684; PubMed=1567844;
RA Omichinski J.G., Clore G.M., Robien M., Saksaguchi K., Appella E.,
RA Gronenborn A.M.;
RT "High-resolution solution structure of the double Cys2His2 zinc
RT finger from the human enhancer binding protein MBP-1.";
RL Biochemistry 31:3907-3917(1992).
CC -1- FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE
CC 5'-GGGACTTCC-3' WHICH IS FOUND IN THE ENHANCER ELEMENTS OF
CC NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1.
CC IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS
CC OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I
CC MGC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT
CC IN T-CELL ACTIVATION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- INDUCTION: BY MITOGEN AND PHORBOL ESTER.
CC -1- DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY
CC SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH
CC ZINC-FINGER IN-BETWEEN.
CC -1- SIMILARITY: STRONG, TO HIVBP2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X51435; CA35798.1; -;
DR PIR; A34203; A34203.
DR PDB; 3ZNF; 15-JAN-92.
DR PDB; 4ZNF; 15-JAN-92.
DR PDB; 1BBO; 31-OCT-93.
DR TRASPAC; T00497; -;
DR GeneW; HGNC:4920; HIVBP1.
DR MIM; 194540; -;
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0003677; F:DNA binding activity; TAS.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 5.

DR SMART; SM00355; ZNF_C2H2; 4.
 DR PROSITE; PS00026; ZINC_FINGER_C2H2_1; 4.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
 KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
 KM Nuclear protein; Repeat; 3D-structure.
 FT ZN_FING 406 428 C2H2-TYPE.
 FT ZN_FING 434 456 C2H2-TYPE.
 FT ZN_FING 958 981 C2HC-TYPE (POTENTIAL).
 FT ZN_FING 2087 2109 C2H2-TYPE.
 FT ZN_FING 2115 2139 C2H2-TYPE.
 FT DOMAIN 803 806 POLY-SER.
 FT STRAND 2088 2088
 FT TURN 2090 2092
 FT STRAND 2095 2095
 FT HELIX 2099 2108
 FT TURN 2109 2109
 FT STRAND 2115 2116
 FT STRAND 2123 2124
 FT HELIX 2127 2135
 SO SEQUENCE 2717 AA; 292217 MW; D4SD3CA951FEA561 CRC64;
 Query Match 66.7%; Score 32; DB 1; Length 2717;
 Best Local Similarity 66.7%; Pred. No. 1.8e+02;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 3 VVPXGXSYS 11
 DB 2405 VVPAGLTVS 2413

RESULT 11
 YHAI_CRYPA STANDARD; PRT; 319 AA.
 ID_YHAI_CRYPA STANDARD; PRT; 319 AA.
 AC P10941;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein 1 in hypovirulence-associated ds-RNA genetic
 DE element (Contains: P29 proteinase).
 OS Cryphonectria parasitica (Chestnut blight fungus) (Endothia
 OS parasitica).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetes; Diaporthales; Valsaceae;
 OC Cryphonectria-Endothia complex; Cryphonectria.
 OX NCBI_Taxid=5116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP713;
 RX MEDLINE=69251594; PubMed=2721496;
 RA Rae B.P., Hillman B.I., Tartaglia J., Nuss D.L.;
 RT "Characterization of double-stranded RNA genetic elements associated
 RT with biological control of chestnut blight: organization of terminal
 RT domain and identification of gene products.";
 RL EMO J. 8:657-663(1989).
 CC -1- MISCELLANEOUS: DOUBLE-STRANDED RNA GENETIC ELEMENTS ARE ASSOCIATED
 CC WITH BIOLOGICAL CONTROL OF THE FUNGAL DISEASE CHESTNUT BLIGHT.
 CC THIS DS-RNA ARE ASSOCIATED WITH HYPOVIRULENCE. THEY ARE LOCALIZED
 CC IN THE CYTOPLASM.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C7.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X15524; CAJ32666.1; -.
 DR PIR; S03833; S03833.
 DR MEROPS; C07.001; -.
 DR InterPro; IPR002704; Peptidase_C7.
 DR Pfam; PF01830; Peptidase_C7; 1.

DR ProDom; PD040949; Peptidase_C7; 1.
 KW Hypothetical protein; Hydrolase; Thiol protease.
 SO SEQUENCE 319 AA; 35443 MW; A1F5775F9A8C7A CRC64;
 Query Match 64.6%; Score 31; DB 1; Length 319;
 Best Local Similarity 85.7%; Pred. No. 32;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 BEVVPXG 7
 DB 31 BEVVPAG 37

RESULT 12
 RPBB_SALTY STANDARD; PRT; 361 AA.
 ID_RPBB_SALTY STANDARD; PRT; 361 AA.
 AC P26391;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE dTDP-glucose 4,6-dehydratase (BC 4.2.1.46).
 GN RPBB OR STM2097.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_Taxid=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2;
 RX MEDLINE=91260454; PubMed=1710759;
 RA Jiang X.-M., Neal B., Santiago F., Lee S.J., Romana L.K., Reeves P.R.;
 RT "Structure and sequence of the rfb (O antigen) gene cluster of
 RT Salmonella serovar typhimurium (strain LT2)." ;
 RL Mol. Microbiol. 5:695-713(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SCS8C1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.B., Spleth J., Clifton S.W., Latreille P.,
 RA Courtney L., Potwilk S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney B.,
 RA Ryan B., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2." ;
 RL Nature 413:852-856(2001).
 CC -1- CATALYTIC ACTIVITY: dTDP-glucose = dTDP-4-dehydro-6-deoxy-D-
 CC glucose + H(2)O.
 CC -1- COFACTOR: NAD.
 CC -1- PATHWAY: dTDP-L-RHAMNOSE BIOSYNTHESIS WITHIN THE O ANTIGEN
 CC BIOSYNTHESIS PATHWAY OF LIPOPOLYSACCHARIDE BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE SUGAR EPIMERASE FAMILY. dTDP-GLUCOSE
 CC DEHYDRATASE SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X56793; CAJ40115.1; -.
 DR EMBL; AB008792; AAL21001.1; -.
 DR PIR; S15299; S15299.
 DR PDB; 1GAL; 21-MAR-01.
 DR PDB; 1KEU; 25-JAN-02.
 DR PDB; 1KEW; 25-JAN-02.
 DR StGene; SG10345; rfbB.
 DR InterPro; IPR005888; dTDP-gluc_dehyd.
 DR InterPro; IPR001509; Epimerase_Dh.
 DR Pfam; PF01370; Epimerase; 1.
 DR TIGRfam; TIGR01181; dTDP-gluc_dehyd; 1.

KM Lipopolysaccharide biosynthesis; Lyase; NAD; Complete proteome;

PT NP_BIND 7 13 NAD (POTENTIAL).
SQ SEQUENCE 361 AA; 40718 MW; 3A574B4D917BC57 CRC64;

Query Match 64.6%; Score 31; DB 1; Length 361;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 EHVVPKXGSYS 10
Db 278 DEIVPKATSY 287

RESULT 13
THD1_BURCE STANDARD; PRT; 507 AA.

AC P53607;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Threonine dehydratase biosynthetic (EC 4.3.1.19) (Threonine deaminase).
GN ILVA.
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OX NCBI_TaxID=292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17616;
RA Bartell J.B., Lesie T.G.;
RU Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: Catalyzes the formation of alpha-ketobutyrate from
threonine in a two-step reaction. The first step is a dehydration
of threonine, followed by rehydration and liberation of ammonia.
CC -1- CATALYTIC ACTIVITY: L-threonine = 2-oxobutanoate + NH(3).
CC -1- COFACTOR: Pyridoxal phosphate.
CC -1- PATHWAY: Isoleucine biosynthesis; first step.
CC -1- SUBUNIT: Homotetramer (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; U40630; AAA83215.1; -.
DR HSSP; P04968; ITDJ.
DR InterPro; IPR001926; B6 enzyme beta.
DR InterPro; IPR000634; S/T_dehydratase.
DR InterPro; IPR005787; Thr_dehydratase.
DR InterPro; IPR001721; ThrDh_C.
DR Pfam; PF00291; PALP; 1.
DR Pfam; PF00585; Thr_dehydrat_C; 2.
DR TIGRFAMs; TIGR01124; ILVA_2Cterm; 1.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
KW Isoleucine biosynthesis; Lyase; Pyridoxal phosphate.
FT BINDING 52
FT SEQUENCE 507 AA; 55326 MW; E9A5D110B0597664 CRC64;

Query Match 64.6%; Score 31; DB 1; Length 507;
Best Local Similarity 60.0%; Pred. No. 51;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 EHVVPKXGSYS 11
Db 120 EVI0AGESYS 129

RESULT 14
SYL_LACLA STANDARD; PRT; 829 AA.

AC Q9CH86;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (leucine--tRNA ligase) (leuRS).
GN LEUS OR IL0816.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Winkler P., Manger S., Jallion O., Malarne K.,
RA Weisenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis ssp. lactis IL1403.";
RU Genome Res. 11:731-753(2001).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(leu) = AMP +
diphosphate + L-leucyl-tRNA(leu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; AE006315; AAC04914.1; -.
DR PIR; H86726; H86726.
DR HAMAP; MF_00049; -1.
DR InterPro; IPR002302; Leu-tRNA-synt1a.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001411; tRNA-synt_1.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR TIGRFAMs; TIGR00396; leuS bact; 1.
DR PROSITE; PS00178; AA tRNA_LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KM Complete proteome.
FT SITE 40
FT SITE 51 "HIGH" REGION.
FT SITE 609 613 "KMSKS" REGION.
FT BINDING 612 612 ATP (By similarity).
FT SEQUENCE 829 AA; 93791 MW; F0F0A43014F6C389 CRC64;

Query Match 64.6%; Score 31; DB 1; Length 829;
Best Local Similarity 66.7%; Pred. No. 85;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EHVVPKXGSYS 9
Db 168 EHVLPDSTS 176

RESULT 15
SYL_SCHPO STANDARD; PRT; 877 AA.

AC Q7A377;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable sulfate permease C3H7.02.

GN SPC3H7.02.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.

```

OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
EX MEDLINE=21846401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leacher S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeil C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkhardt G., Aert R., Robben J., Grymompres B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moesl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wandut R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerruti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE INTO THE CELL (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE SLC26A FAMILY OF TRANSPORTERS.
CC -!- SIMILARITY: Contains 1 STAS domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, AL031261; CNA20298.1; -.
DR PIR, T40413; T40413.
DR GenDB, SPombe; SPBC3H7.02; -.
DR InterPro; IPR002645; STAS.
DR InterPro; IPR001902; Sulph_transp.
DR Pfam; PF01740; STAS; 1.
DR Pfam; PF00916; Sulfate_transp; 1.
DR TIGRPFAMs; TIGR00815; sulP; 1.
DR PROSITE; PS01130; SLC26A; 1.
DR PROSITE; PS50801; STAS; 1.
KW Transport; Transmembrane.
FT TRANSMEM 133 153 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 186 206 POTENTIAL.
FT TRANSMEM 221 241 POTENTIAL.
FT TRANSMEM 243 263 POTENTIAL.
FT TRANSMEM 292 312 POTENTIAL.
FT TRANSMEM 329 349 POTENTIAL.
FT TRANSMEM 384 404 POTENTIAL.
FT TRANSMEM 424 444 POTENTIAL.
FT TRANSMEM 461 481 POTENTIAL.
FT TRANSMEM 484 504 POTENTIAL.
FT TRANSMEM 518 538 POTENTIAL.
FT TRANSMEM 543 563 POTENTIAL.
FT DOMAIN 594 747 STAS.
SQ SEQUENCE 877 AA; 96373 MW; 56995A8493371E43 CRC64;

```

Query Match

64.6%; Score 31; DB 1; Length 877;

Best Local Similarity 66.7%; Pred. No. 91;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VPEXGXSYS 11
DB 148 VPEGMSYA 156

Search completed: December 22, 2003, 17:42:28
Job time : 4.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:27:26 ; Search time 25.2 Seconds

(without alignments)
112.642 Million cell updates/sec

Title: US-09-909-164-43
Perfect score: 48
Sequence: 1 EHVVPXGXSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_todent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriopl:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	77.1	471	11	Q8R126
2	37	77.1	484	11	Q8VD18
3	37	77.1	484	11	Q8BTX4
4	37	77.1	484	11	Q8BK35
5	37	77.1	1044	16	Q8D1H0
6	36	75.0	3472	1	Q74056
7	35	72.9	225	10	Q40129
8	34	70.8	143	17	Q8TX62
9	34	70.8	174	10	Q9M374
10	34	70.8	479	4	Q96CS0
11	34	70.8	541	16	Q98BBS
12	34	70.8	678	16	Q9R1X6
13	34	70.8	1028	16	Q8YJ11
14	34	70.8	1123	16	Q8EMD4
15	34	70.8	1442	17	Q96YH5
16	33	68.8	78	6	Q9XST4

17	33	68.8	175	6	Q81033	Q81033 bos taurus
18	33	68.8	200	16	Q8XHX3	Q8XHX3 clostridium
19	33	68.8	215	6	Q81031	Q81031 bos taurus
20	33	68.8	217	4	Q00404	Q00404 homo sapien
21	33	68.8	297	6	Q8HX19	Q8HX19 bos taurus
22	33	68.8	299	4	Q9UER9	Q9UER9 homo sapien
23	33	68.8	495	11	Q8CID7	Q8CID7 mus musculu
24	33	68.8	526	5	Q9VMN9	Q9VMN9 drosophila
25	33	68.8	815	10	Q9SP93	Q9SP93 arabidopsis
26	33	68.8	815	10	Q8L850	Q8L850 arabidopsis
27	33	68.8	899	16	Q8G415	Q8G415 bifidobacte
28	33	68.8	933	5	Q8SS39	Q8SS39 encaphalito
29	33	68.8	1237	16	Q8EFM1	Q8EFM1 shewanella
30	32	66.7	96	3	Q9R855	Q9R855 kuyveromyc
31	32	66.7	219	17	Q971S2	Q971S2 sulfolobus
32	32	66.7	227	16	Q9RZU8	Q9RZU8 deinococcus
33	32	66.7	290	4	Q96WU1	Q96WU1 homo sapien
34	32	66.7	387	16	Q98FX1	Q98FX1 rhizobium 1
35	32	66.7	387	16	Q92MD6	Q92MD6 rhizobium m
36	32	66.7	413	11	Q8K289	Q8K289 mus musculu
37	32	66.7	425	5	Q9XYK4	Q9XYK4 caenorhabdi
38	32	66.7	489	4	Q81YK3	Q81YK3 homo sapien
39	32	66.7	556	4	Q43733	Q43733 homo sapien
40	32	66.7	583	5	Q9BH63	Q9BH63 plasmodium
41	32	66.7	583	5	Q9BHA5	Q9BHA5 plasmodium
42	32	66.7	583	5	Q81S67	Q81S67 plasmodium
43	32	66.7	602	10	Q9LUB0	Q9LUB0 arabidopsis
44	32	66.7	630	2	Q87110	Q87110 comamonas a
45	32	66.7	670	11	Q01487	Q01487 rattus ratt

ALIGNMENTS

RESULT 1

Q8R126 PRELIMINARY; PRT; 471 AA.

AC Q8R126.

DT 01-JUN-2002 (TREMBLrel. 21, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DB Hypothetical 54.5 kDa protein (Fragment).

GN GLTSCR2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

RA Strauberg R.;

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC025810; AAH25810.1; -

DR MGD; MGI:2154441; Glutcr2.

KW Hypothetical protein.

FT NON TER 1

SO SEQUENCE 471 AA; 54506 MW; E0DA685C374A9760 CRC64;

Query Match 77.1%; Score 37; DB 11; Length 471;

Best Local Similarity 60.0%; Pred. No. 12;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

RESULT 2

Q8VD18 PRELIMINARY; PRT; 484 AA.

AC Q8VD18.

DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DS Similar to glioma tumor suppressor candidate region gene 2.
 GN GLUTSCR2 OR AW536441.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Salivary gland;
 RA Strausberg R.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC017637; AAH17637.1; --
 DR MGI; MGI:2154441; Gtsc2.
 SQ SEQUENCE 484 AA; 55835 MW; BBB45F3B4BE02A36 CRC64;

 Query Match 77.1%; Score 37; DB 11; Length 484;
 Best Local Similarity 60.0%; Pred. No. 13;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

 QY 2 EVVFXGXSYS 11
 DB 239 EVIPAGASYN 248

 RESULT 3
 Q8BTX4 PRELIMINARY; PRT; 484 AA.
 AC Q8BTX4;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Similar to glioma tumor suppressor CANDIDATE region gene 2
 DE protein.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NOJ; TISSUE=Thymus;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK088461; BAC40367.1; --
 SQ SEQUENCE 484 AA; 55806 MW; B3056425B5EBCAD8 CRC64;

 Query Match 77.1%; Score 37; DB 11; Length 484;
 Best Local Similarity 60.0%; Pred. No. 13;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

 QY 2 EVVFXGXSYS 11
 DB 239 EVIPAGASYN 248

 RESULT 4
 Q8BK35 PRELIMINARY; PRT; 484 AA.
 AC Q8BK35;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Similar to glioma tumor suppressor CANDIDATE region gene 2
 DE protein.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Pituitary;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK077341; BAC6760.1; --
 SQ SEQUENCE 484 AA; 55792 MW; BB67949BCB92D44 CRC64;

 Query Match 77.1%; Score 37; DB 11; Length 484;
 Best Local Similarity 60.0%; Pred. No. 13;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

 QY 2 EVVFXGXSYS 11
 DB 239 EVIPAGASYN 248

 RESULT 5
 Q8DIH0 PRELIMINARY; PRT; 1044 AA.
 ID Q8DIH0
 AC Q8DIH0;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Multidrug efflux transporter.
 GN TLL1618.
 OS Synecococcus elongatus (Thermosynechococcus elongatus).
 CC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
 CX NCBI_TaxID=32046;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP-1;
 RX MEDLINE=2222144; PubMed=12240834;
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A., Nakazaki N.,
 RA Shimpō S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the thermophilic cyanobacterium
 RT Thermosynechococcus elongatus BP-1."
 RL DNA Res. 9:123-130(2002).
 DR EMBL; AP005374; BAC09170.1; --
 KW Complete proteome.
 SQ SEQUENCE 1044 AA; 113205 MW; 00B9C13F0F636D2F CRC64;

 Query Match 77.1%; Score 37; DB 16; Length 1044;
 Best Local Similarity 63.6%; Pred. No. 30;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

 QY 1 EVVFXGXSYS 11
 DB 843 EVLPNGIGYS 853

 RESULT 6
 O74056 PRELIMINARY; PRT; 3472 AA.
 ID O74056
 AC O74056;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DB Hypothetical 367.1 kDa protein.
 OS Cenarchaeum symbiosum.
 CC Archaea; Crenarchaeota; Thermoprotei; Cenarchaeales; Cenarchaeaceae;
 CC Cenarchaeum.
 CX NCBI_TaxID=46770;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B;
 RX MEDLINE=98422450; PubMed=9748430;
 RA Schleper C., DeLong E.F., Preston C.M., Feldman R.A., Wu K.Y.,

```

RA Swanson R.V.;
RT "Genomic analysis reveals chromosomal variation in natural populations
RT of the uncultured psychrophilic archaeon Cenarchaeum symbiosum.";
RL J. Bacteriol. 180:5003-5009 (1998).
DR EMBL; AF083072; AAC62699.1; -.
DR InterPro; IPR000515; BDP_transp.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 4.
DR SMART; SM00320; WD40; 2.
DR PROSITE; PS00402; BDP_TRANSP_INN_MEMBER; 1.
KM Hypothetical protein; Repeat; WD repeat.
SQ SEQUENCE 3472 AA; 367058 MW; 37F80707030F9355 CRC64;

Query Match
Best Local Similarity 54.5%; Score 36; DB 1; Length 3472;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGXSYS 11
DB 2294 EDVIPRGISFS 2304

RESULT 7
Q40129 PRELIMINARY; PRT; 225 AA.
ID Q40129;
AC Q40129;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical 25.2 kDa protein precursor.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
CX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF36; Tissue=Platell;
RX MEDLINE=95375233; PubMed=7647301;
RA Milligan S.B., Gaseer C.S.;
RT "Nature and regulation of plectin-expressed genes in tomato.";
RL Plant Mol. Biol. 28:691-711 (1995).
DR EMBL; U20592; AAA0497.1; -.
DR InterPro; IPR002160; Kunitz_legume.
DR Pfam; PF0197; Kunitz_legume; 1.
DR PRINTS; PR00291; KUNITZINSTR.
DR ProDom; PD000891; Kunitz_legume; 1.
DR SMART; SM00452; STI; 1.
DR PROSITE; PS00283; SOYBEAN_KUNITZ; 1.
KM Hypothetical protein; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 225 UNKNOWN.
SQ SEQUENCE 225 AA; 25188 MW; 1074C261D20CEDAD CRC64;

Query Match
Best Local Similarity 54.5%; Score 35; DB 10; Length 225;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGXSYS 11
DB 32 DEVVNGKTYA 42

RESULT 8
Q8TX62 PRELIMINARY; PRT; 143 AA.
ID Q8TX62;
AC Q8TX62;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Uncharacterized conserved protein.
GN MK0814.
OS Mechanopyrus kandleri.

```

```

OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Mechanopyrus.
CX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
RA Shcherbina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natarle D.A., Rogozhin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Mal'khin A.G., Koonin E.V., Kozlovskiy S.A.;
RT "The complete genome of hyperthermophilic Methanopyrus kandleri AV19
RT and monophyly of archaeal mechanogens.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).
DR EMBL; AB010372; AAM02027.1; -.
DR InterPro; IPR001602; UPF0047.
DR Pfam; PF01894; UPF0047; 1.
DR ProDom; PD005232; UPF0047; 1.
DR TIGRPFAMs; TIGR00149; TIGR00149; 1.
DR PROSITE; PS01314; UPF0047; 1.
KM Complete proteome.
SQ SEQUENCE 143 AA; 15734 MW; 4C8B28A1FB8DD0B CRC64;

Query Match
Best Local Similarity 70.8%; Score 34; DB 17; Length 143;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXSYS 10
DB 75 BEVVPQAGY 84

RESULT 9
Q9M374 PRELIMINARY; PRT; 174 AA.
ID Q9M374;
AC Q9M374;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Acidic endochitinase (EC 3.2.1.14) (Fragment).
GN P3A.
OS Betula verrucosa (White birch) (Betula pendula).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Euxoide I; Fagales; Betulaceae; Betula.
CX NCBI_TaxID=3505;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RA Hillovaara-Teljo M., Korhonen M.S., Pajva T.E., Kangasjarvi J.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ279692; CAB66334.1; -.
DR HSBP; P23472; ZHM.
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR001579; Glyco_hydro_18/2.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR PROSITE; PS01095; CHITINASE_18; 1.
KM Glycosidase; Hydrolase.
FT NON_TER 1 174
FT NON_TER 174 174
SQ SEQUENCE 174 AA; 17936 MW; 834ADC6B5C76634 CRC64;

Query Match
Best Local Similarity 70.8%; Score 34; DB 10; Length 174;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXSYS 11
DB 74 VVPQGSYS 82

RESULT 10
Q96CS0

```

ID 096CS0 PRELIMINARY; PRT; 479 AA.
 AC 096CS0; 096IT7;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 22, Last annotation update)
 DE Gloma tumor suppressor candidate region gene 2 (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=lung, and Skin;
 RA Straube R.;
 RL Submitted (SSP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC014009; AA014009.1; -
 DR EMBL; BC007248; AA07248.1; -
 FT NON TER 1
 SO SEQUENCE 479 AA; 54529 MW; 0A6C3A8B476F7E8F CRC64;

Query Match 70.8%; Score 34; DB 4; Length 479;
 Best Local Similarity 60.0%; Pred. No. 59;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVVPXGXSYS 11
 DB 240 EVAPAGASYN 249

RESULT 11

ID 098BP5 PRELIMINARY; PRT; 541 AA.
 AC 098BP5;
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Probable DNA ligase.
 OS ML5481.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF30309; PubMed=11214968;
 RX MEDLINE=21082930;
 RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Matenabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti."
 RT Mesorhizobium loti.
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP003006; BAB51927.1; -
 DR InterPro; IPR000977; DNA_ligase.
 DR Pfam; PF01068; DNA_ligase_A; 1.
 DR Pfam; PF04679; DNA_ligase_A; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A; 1.
 KW Complete proteome.
 SO SEQUENCE 541 AA; 60645 MW; 2BEBF705453F28F8 CRC64;

Query Match 70.8%; Score 34; DB 16; Length 541;
 Best Local Similarity 60.0%; Pred. No. 67;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGXSYS 10
 DB 445 BELVPGKAY 454

RESULT 12

ID 09E1X6 PRELIMINARY; PRT; 678 AA.
 AC 09E1X6;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Hypothetical 75.9 kDa protein.
 OS Cercopithecine herpesvirus 7.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicelloviruses.
 OX NCBI_TaxID=35245;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Gray W.L., Starnes H.B., White M.W., Ashburn C.V., Mahalingam R.;
 RA "Complete sequence of the Simian Varicella Virus Genome."
 RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF275348; AA027217.1; -
 DR EMBL; AF275348; AA027217.1; -
 KW Hypothetical protein.
 SO SEQUENCE 678 AA; 75850 MW; A17B09B30512F23C CRC64;

Query Match 70.8%; Score 34; DB 12; Length 678;
 Best Local Similarity 50.0%; Pred. No. 86;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXSYS 10
 DB 147 BEIIPKSTRY 156

RESULT 13

ID 08YJ11 PRELIMINARY; PRT; 1028 AA.
 AC 08YJ11;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE ATP-dependent DNA helicase.
 OS BMR1025.
 OS Brucella melitensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29459;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=16M / ATCC 23456 / Biotype 1;
 RX MEDLINE=20020109; PubMed=11756688;
 RA DelVecchio V.G., Kapral V., Redkar R.J., Patra G., Mujar C., Los T.,
 RA Ivanova N., Anderson I., Bhattacharya A., Lyridis A., Reznik G.,
 RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Coleman E.,
 RA Selkov E., Bizer P.H., Hagius S., O'Callaghan D., Leiserson J.-J.,
 RA Haselhorn R., Kyriades N., Overbeek R.;
 RT "The genome sequence of the facultative intracellular pathogen
 Brucella melitensis."
 RT Brucella melitensis.
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 DR EMBL; AB009470; AA051457.1; -
 DR InterPro; IPR002086; Aldehyde_dehydr.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR001650; Helicase_C.
 DR Pfam; PF00271; Helicase_C; 1.
 DR SMART; SM00490; HELIC_C; 1.
 DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Complete proteome.
 SO SEQUENCE 1028 AA; 112996 MW; A752B70425728219 CRC64;

Query Match 70.8%; Score 34; DB 16; Length 1028;
 Best Local Similarity 54.5%; Pred. No. 1.4e+02;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXSYS 11
 DB 76 BKIVPGARYS 86

Query Match	70.8%	Score 34	DB 16	Length 1123
Best Local Similarity	70.0%	Pred. No. 1.5e+02		
Matches	7	Conservative	0	Mismatches 3
				Indels 0
				Gaps 0
QY	2	EVVPRKGSYS	11	
Db	658	EYVPMGLSYS	667	

Query Match	70.8%	Score	34	DB	17	Length	1442
Best Local Similarity	50.0%	Pred. No.	2e+02				
Matches	5	Conservative	2	Mismatches	3	Indels	0
						Gaps	0

Qy	1	EEVVPXGXS	10
	:	:	
Db	863	EEITPTGANY	872

CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX

Sequence 11 AA;

Query Match 94.2%; Score 49; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0037;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EKVVPXGSHYS 11
DB 1 EKVVPXGSHYS 11

RESULT 4
ABB80555 ID ABB80555 standard; peptide; 11 AA.

AC ABB80555;
DT 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #35.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virocidine.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

XX W0200208251-A2.

PD 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX LAm-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virocidine activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

Sequence 11 AA;

Query Match 94.2%; Score 49; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0037;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EKVVPXGSHYS 11
DB 1 EKVVPXGSHYS 11

RESULT 5

ABB80523 ID ABB80523 standard; peptide; 11 AA.

AC ABB80523;

DT 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #3.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virocidine.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

XX W0200208251-A2.

PD 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX LAm-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virocidine activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

Sequence 11 AA;

Query Match 92.3%; Score 48; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0058;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EKVVPXGSHYS 11
DB 1 EKVVPXGSHYS 11

DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #11.
XX	
KM	Hepatitis C virus; HCV, serine protease; inhibitor; alpha-ketoamide;
KW	virucide.
XX	
OS	Synthetic.
XX	
PH	Key
FT	Modified-site
FT	/note= "N-terminal acetyl"
FT	6
FT	/note= "Normalyl carbonyl forming keto-amide linkage with
FT	residue 7"
FT	Modified-site
FT	11
FT	/note= "C-terminal amide"
XX	
FN	WO200208251-A2.
XX	
PD	31-JAN-2002.
XX	
PP	19-JUL-2001, 2001WO-US23169.
XX	
PR	21-JUL-2000; 2000US-220101P.
XX	
PA	(CORV-) CORVAS INT INC.
XX	
PI	Llm-wilby M, Levy OE, Brunck TK,
XX	
DR	WPI; 2002-361643/39.
XX	
PT	Novel peptide compound having hepatitis C virus protease inhibitory
FT	activity useful for treating disorders associated with hepatitis C
FT	virus protease
XX	
PS	Claim 17; Page 64; 69pp; English.
XX	
CC	The sequence represents a peptide compound of the invention having
CC	hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC	invention are alpha-ketoamide peptide analogues. The peptides have
CC	virucide activity, and are useful for treating and in the manufacture of
CC	a pharmaceutical to treat disorders associated with HCV protease. A
CC	pharmaceutical composition comprising the peptide as an active ingredient
CC	is useful for treating disorders associated with hepatitis C virus.
XX	
SC	Sequence 11 AA;
XX	
Query Match	92.3%; Score 48; DB 23; Length 11;
Best Local Similarity	90.9%; Pred. No. 0.0058;
Matches 10; Conservative	0; Mismatches 1; Indels 0; Gaps 0.
OY	1 EEVVPXGHHYS 11
DB	1 EEVVPXGGHYS 11
XX	
RESULT 8	
ID	ABB80532 standard; peptide; 11 AA.
XX	
AC	ABB80532;
XX	
DT	08-OCT-2002 (first entry)
XX	
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #12.
XX	
KW	Hepatitis C virus; HCV, serine protease; inhibitor; alpha-ketoamide;
KM	virucide.
XX	
OS	Synthetic.
XX	
PH	Key
FT	Modified-site
FT	Location/Qualifiers
FT	1
FT	/note= "N-terminal acetyl"
FT	6
FT	/note= "Normalyl carbonyl forming keto-amide linkage with
FT	residue 7"
FT	Modified-site
FT	11
FT	/note= "C-terminal amide"
XX	
FN	WO200208251-A2.
XX	
PD	31-JAN-2002.
XX	
PP	19-JUL-2001, 2001WO-US23169.
XX	
PR	21-JUL-2000; 2000US-220101P.
XX	
PA	(CORV-) CORVAS INT INC.
XX	
PI	Llm-wilby M, Levy OE, Brunck TK,
XX	
DR	WPI; 2002-361643/39.
XX	
PT	Novel peptide compound having hepatitis C virus protease inhibitory
FT	activity useful for treating disorders associated with hepatitis C
FT	virus protease
XX	
PS	Claim 17; Page 64; 69pp; English.
XX	
CC	The sequence represents a peptide compound of the invention having
CC	hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC	invention are alpha-ketoamide peptide analogues. The peptides have
CC	virucide activity, and are useful for treating and in the manufacture of
CC	a pharmaceutical to treat disorders associated with HCV protease. A
CC	pharmaceutical composition comprising the peptide as an active ingredient
CC	is useful for treating disorders associated with hepatitis C virus.
XX	
SC	Sequence 11 AA;
XX	
Query Match	92.3%; Score 48; DB 23; Length 11;
Best Local Similarity	90.9%; Pred. No. 0.0058;
Matches 10; Conservative	0; Mismatches 1; Indels 0; Gaps 0.
OY	1 EEVVPXGHHYS 11
DB	1 EEVVPXGGHYS 11
XX	
RESULT 8	
ID	ABB80532 standard; peptide; 11 AA.
XX	
AC	ABB80532;
XX	
DT	08-OCT-2002 (first entry)
XX	
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #12.
XX	
KW	Hepatitis C virus; HCV, serine protease; inhibitor; alpha-ketoamide;
KM	virucide.
XX	
OS	Synthetic.
XX	
PH	Key
FT	Modified-site
FT	Location/Qualifiers
FT	1
FT	/note= "N-terminal acetyl"
FT	6
FT	/note= "Normalyl carbonyl forming keto-amide linkage with
FT	residue 7"
FT	Modified-site
FT	11
FT	/note= "C-terminal amide"
XX	
FN	WO200208251-A2.
XX	
PD	31-JAN-2002.
XX	
PP	19-JUL-2001, 2001WO-US23169.
XX	
PR	21-JUL-2000; 2000US-220101P.
XX	
PA	(CORV-) CORVAS INT INC.
XX	
PI	Llm-wilby M, Levy OE, Brunck TK,
XX	
DR	WPI; 2002-361643/39.
XX	
PT	Novel peptide compound having hepatitis C virus protease inhibitory
FT	activity useful for treating disorders associated with hepatitis C
FT	virus protease
XX	
PS	Claim 17; Page 64; 69pp; English.
XX	
CC	The sequence represents a peptide compound of the invention having
CC	hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC	invention are alpha-ketoamide peptide analogues. The peptides have
CC	virucide activity, and are useful for treating and in the manufacture of
CC	a pharmaceutical to treat disorders associated with HCV protease. A
CC	pharmaceutical composition comprising the peptide as an active ingredient
CC	is useful for treating disorders associated with hepatitis C virus.
XX	
SC	Sequence 11 AA;
XX	
Query Match	92.3%; Score 48; DB 23; Length 11;
Best Local Similarity	90.9%; Pred. No. 0.0058;
Matches 10; Conservative	0; Mismatches 1; Indels 0; Gaps 0.
OY	1 EEVVPXGHHYS 11
DB	1 EEVVPXGGHYS 11
XX	

FT Modified-site /note= "N-terminal acetyl"
FT 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT WO200208251-A2.
PN 31-JAN-2002.
PD 19-JUL-2001; 2001WO-US23169.
PE 21-JUL-2000; 2000US-220101P.
PR (CORV-) CORVAS INT INC.
PA Lim-wilby M, Levy OE, Brunck TK;
PI WPI; 2002-361643/39.
DR
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 92.3%; Score 48; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0058;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGSHYS 11
DB 1 EEVVPXGSHYS 11
RESULT 9
ABB80537
ID ABB80537 standard; peptide; 11 AA.
XX
AC ABB80537;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #17.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT

XX
PN WO200208251-A2.
XX 31-JAN-2002.
PD 19-JUL-2001; 2001WO-US23169.
PE 21-JUL-2000; 2000US-220101P.
PR (CORV-) CORVAS INT INC.
PA Lim-wilby M, Levy OE, Brunck TK;
PI WPI; 2002-361643/39.
DR
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 92.3%; Score 48; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0058;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGSHYS 11
DB 1 EEVVPXGSHYS 11
RESULT 10
ABB80541
ID ABB80541 standard; peptide; 11 AA.
XX
AC ABB80541;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #21.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
PN WO200208251-A2.
PD 31-JAN-2002.
PE 19-JUL-2001; 2001WO-US23169.
PF 21-JUL-2000; 2000US-220101P.
PR

XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 92.3%; Score 48; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.0058;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11
 |||||
 DB 1 EEVVPXGXHYS 11

RESULT 13

ABB80544
 ID ABB80544 standard; peptide; 11 AA.

AC ABB80544;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #24.

KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KM virucide.

OS Synthetic.

Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"

FT Modified-site 11 /note= "C-terminal amide"

PN W0200208251-A2.

PD 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

PA (CORV-) CORVAS INT INC.

PI Lim-wilby M, Levy OB, Brunck TK;

DR WPI; 2002-361643/39.

PT Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -

PS Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 76.9%; Score 40; DB 23; Length 11;
 Best Local Similarity 81.8%; Pred. No. 0.23;
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11
 |||||
 DB 1 EEVVPXGXHYS 11

RESULT 14

ABB80545
 ID ABB80545 standard; peptide; 11 AA.

AC ABB80545;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #25.

KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KM virucide.

OS Synthetic.

Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

PN W0200208251-A2.

PD 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

PA (CORV-) CORVAS INT INC.

PI Lim-wilby M, Levy OB, Brunck TK;

DR WPI; 2002-361643/39.

PT Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -

PS Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 76.9%; Score 40; DB 23; Length 11;
 Best Local Similarity 81.8%; Pred. No. 0.23;
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXHYS 11
 |||||
 DB 1 EEVVPXGXHYS 11

RESULT 15

ABB80547 standard; peptide; 11 AA.

AC ABB80547;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #27.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketamide; virucide.

OS Synthetic.

FH Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"

FT Modified-site 11 /note= "C-terminal amide"

PN WO200208251-A2.

PD 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

PA (CORV-) CORVAS INT INC.

PI Lim-wilby M, Levy OB, Brunck TK;

DR WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory

PT activity useful for treating disorders associated with hepatitis C

PS Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having

CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

CC invention are alpha-ketamide peptide analogues. The peptides have

CC virucide activity, and are useful for treating and in the manufacture of

CC a medicament to treat disorders associated with HCV protease. A

CC pharmaceutical composition comprising the peptide as an active ingredient

CC is useful for treating disorders associated with hepatitis C virus.

SQ Sequence 11 AA;

OY 1 BEVVPXGXHS 11

DB 1 BEVVPXGTDYS 11

Search completed: December 22, 2003, 17:41:02

Job time : 32.4667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using SW model

Run on: December 22, 2003, 16:37:03 / Search time 10.133 Seconds
(without alignments)
45.930 Million cell updates/sec

Title: US-09-909-164-44
Perfect score: 52
Sequence: 1 EAVVPXGKHYS 11

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents AA:

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTCDS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	73.1	856	4 US-09-252-991A-21444	Sequence 21444, A
2	36	69.2	1037	4 US-09-134-001C-4794	Sequence 4794, Ap
3	35	67.3	277	4 US-09-252-991A-26615	Sequence 26615, A
4	34	65.4	126	2 US-08-879-995A-3	Sequence 3, Appli
5	34	65.4	126	3 US-09-215-096-3	Sequence 3, Appli
6	33	63.5	63	4 US-09-107-532A-6771	Sequence 6771, Ap
7	33	63.5	447	4 US-08-961-083-182	Sequence 182, App
8	33	63.5	447	4 US-09-536-784-182	Sequence 6, Appli
9	33	63.5	484	4 US-09-468-656A-6	Sequence 19, Appli
10	33	63.5	600	2 US-08-821-119-19	Sequence 2, Appli
11	33	63.5	600	2 US-08-821-118-2	Sequence 66, Appli
12	33	63.5	763	3 US-08-961-083-66	Sequence 66, Appli
13	33	63.5	763	3 US-09-536-784-66	Sequence 66, Appli
14	33	63.5	796	3 US-08-961-083-56	Sequence 56, Appli
15	33	63.5	796	3 US-09-536-784-56	Sequence 56, Appli
16	33	63.5	819	4 US-09-468-656A-8	Sequence 8, Appli
17	33	63.5	819	4 US-09-468-656A-10	Sequence 10, Appli
18	33	63.5	838	4 US-09-468-656A-4	Sequence 4, Appli
19	32	61.5	253	4 US-09-328-352-6440	Sequence 6440, Ap
20	32	61.5	738	1 US-08-530-010-3	Sequence 3, Appli
21	32	61.5	738	1 US-08-530-010-5	Sequence 5, Appli
22	32	61.5	738	1 US-08-530-010-7	Sequence 7, Appli
23	32	61.5	738	1 US-08-530-010-9	Sequence 9, Appli
24	32	61.5	738	1 US-08-530-010-11	Sequence 11, Appli
25	32	61.5	738	1 US-08-484-101B-3	Sequence 3, Appli
26	32	61.5	738	2 US-08-484-101B-5	Sequence 5, Appli
27	32	61.5	738	2 US-08-484-101B-7	Sequence 7, Appli

28	32	61.5	738	2 US-08-484-101B-9	Sequence 9, Appli
29	32	61.5	738	2 US-08-484-101B-11	Sequence 11, Appli
30	32	61.5	738	3 US-08-714-524D-3	Sequence 3, Appli
31	32	61.5	738	3 US-08-714-524D-5	Sequence 5, Appli
32	32	61.5	738	3 US-08-714-524D-7	Sequence 7, Appli
33	32	61.5	738	3 US-08-714-524D-9	Sequence 9, Appli
34	32	61.5	738	3 US-08-714-524D-11	Sequence 11, Appli
35	32	61.5	1247	4 US-09-252-991A-32960	Sequence 32960, A
36	31	59.6	70	4 US-09-134-001C-3950	Sequence 3950, Ap
37	31	59.6	237	4 US-09-134-001C-4185	Sequence 4185, Ap
38	31	59.6	252	4 US-09-328-352-7284	Sequence 7284, Ap
39	31	59.6	388	4 US-08-178-257-15	Sequence 15, Appli
40	31	59.6	502	4 US-09-342-647-4	Sequence 4, Appli
41	31	59.6	622	2 US-08-459-146-2	Sequence 2, Appli
42	31	59.6	622	2 US-08-459-065-2	Sequence 2, Appli
43	31	59.6	630	4 US-09-342-647-2	Sequence 2, Appli
44	31	59.6	667	4 US-09-342-647-28	Sequence 28, Appli
45	31	59.6	700	4 US-09-621-816B-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-252-991A-21444
Sequence 21444, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT FILING DATE: 1999-02-18

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 21444

LENGTH: 856

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-21444

Query Match

Best local similarity 73.1%; Score 38; DB 4; Length 856;

Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 1 EAVVPXGKHYS 10

64 EAVVPXGKHYS 73

RESULT 2

US-09-134-001C-4794

Sequence 4794, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4794

LENGTH: 1037

TYPE: PRT

ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4794

Query Match 69.2%; Score 36; DB 4; Length 1037;
Best Local Similarity 63.6%; Pred. No. 47;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXHS 11
DB 199 KEVVSNGLHYS 209

RESULT 3
US-09-252-991A-26615
Sequence 26615, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 26615
LENGTH: 277
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26615

Query Match 67.3%; Score 35; DB 4; Length 277;
Best Local Similarity 63.6%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 BEVVPXGXHS 11
DB 48 BEVPGGHTS 58

RESULT 4
US-08-879-995A-3
Sequence 3, Application US/08879995A
Patent No. 5985606
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Kaser, Matthew R.
TITLE OF INVENTION: HUMAN PREPROTACHYKININ B
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/879,995A
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0326 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 163590
US-08-879-995A-3

Query Match 65.4%; Score 34; DB 2; Length 126;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGXH 9
DB 28 BEVVPXGXH 36

RESULT 5
US-09-215-096-3
Sequence 3, Application US/09215096
Patent No. 6008194
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Kaser, Matthew R.
TITLE OF INVENTION: HUMAN PREPROTACHYKININ B
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/215,096
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/879,995
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0326 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 163590

US-09-215-096-3

Query Match 65.4%; Score 34; DB 3; Length 126;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGXH 9
|:|:|:|:
Db 28 EQVPGGGH 36

RESULT 6
US-09-107-532A-6771

; Sequence 6771, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

; CORRESPONDENCE ADDRESS:

; ADDRESSER: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; CITY: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD-ROM ISO9660

; COMPUTER: PC

; OPERATING SYSTEM: <Unknown>

; SOFTWARE: ASCII

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,532A

; FILING DATE: 30-Jun-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/085,598

; FILING DATE: 14 May 1998

; APPLICATION NUMBER: 60/051571

; FILING DATE: July 2, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Arinello, Pamela Deneke

; REGISTRATION NUMBER: 40,489

; REFERENCE/DOCKET NUMBER: GTC-012

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (781)893-5007

; TELEFAX: (781)893-8277

; INFORMATION FOR SEQ ID NO: 6771:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 63 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: YES

; ORIGINAL SOURCE:

; ORGANISM: Enterococcus faecium

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (8) LOCATION 1..63

; SEQUENCE DESCRIPTION: SEQ ID NO: 6771:

US-09-107-532A-6771

Query Match 63.5%; Score 33; DB 4; Length 63;
Best Local Similarity 45.5%; Pred. No. 8.5;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGXH 11
|:|:|:|:
Db 5 BEIMAFGDHN 15

RESULT 7
US-08-961-083-182

; Sequence 182, Application US/08961083

; Patent No. 6159469

; GENERAL INFORMATION:

; APPLICANT: Choi et. al.

; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

; NUMBER OF SEQUENCES: 452

; CORRESPONDENCE ADDRESS:

; ADDRESSER: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/961,083

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; ATTORNEY/AGENT INFORMATION:

; NAME: Brookes, A. Anders

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB340P2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 182:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 447 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-961-083-182

Query Match 63.5%; Score 33; DB 3; Length 447;
Best Local Similarity 62.5%; Pred. No. 74;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXHY 10
|:|:|:|:
Db 178 IVPHGXHY 185

RESULT 8
US-09-536-784-182

; Sequence 182, Application US/09536784

; Patent No. 6573082

; GENERAL INFORMATION:

; APPLICANT: Choi et. al.

; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

; NUMBER OF SEQUENCES: 452

; CORRESPONDENCE ADDRESS:

; ADDRESSER: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/536,784

; FILING DATE: 30-Oct-1997

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 182:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 182:
US-09-536-784-182

Query Match 63.5%; Score 33; DB 4; Length 447;
Best Local Similarity 62.5%; Pred. No. 74;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VPKGXHY 10
DB 178 IVPHGCHY 185

RESULT 9
US-09-468-656A-6
Sequence 6, Application US/09468656A
Patent No. 6582706
GENERAL INFORMATION:
APPLICANT: Johnson, Leslie S.
APPLICANT: Adamu, John B.
TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
FILE REFERENCE: 469201-444
CURRENT APPLICATION NUMBER: US/09/468,656A
CURRENT FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: 60/113,048
PRIOR FILING DATE: 1998-12-21
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 484
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-468-656A-6

Query Match 63.5%; Score 33; DB 4; Length 484;
Best Local Similarity 62.5%; Pred. No. 81;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VPKGXHY 10
DB 202 IVPHGCHY 209

RESULT 10
US-08-821-119-19
Sequence 19, Application US/08821119
Patent No. 5821104
GENERAL INFORMATION:
APPLICANT: Holm, Kaj Andre
APPLICANT: Raasmussen, Gretel
APPLICANT: Halikier, Torben
APPLICANT: Lehmbeck, Jan
TITLE OF INVENTION: Tripeptidyl Aminopeptidase
NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:
ADDRESSER: No. 58211040 No. 5821104disk of No. 5821104th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,119
FILING DATE: 19-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4107,204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 600 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: Internal
US-08-821-119-19

Query Match 63.5%; Score 33; DB 2; Length 600;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 VPKGXHY 11
DB 31 VPKGXHY 38

RESULT 11
US-08-821-118-2
Sequence 2, Application US/08821118
Patent No. 5989889
GENERAL INFORMATION:
APPLICANT: Rey, Michael
APPLICANT: Golihtely, Elizabeth
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING POLYPEPTIDES
TITLE OF INVENTION: HAVING TRIPEPTIDE AMINOPEPTIDASE
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSER: No. 59898890 No. 5989889disk of No. 5989889th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,118
FILING DATE: 19-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728

Tue Dec 23 09:51:14 2003

```

REFERENCE/DOCKET NUMBER: 4107.400-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 600 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-821-118-2

Query Match      63.5%; Score 33; DB 2; Length 600;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 4 VPKGXHYS 11
 DB 31 VPKGXHYS 38

```

RESULT 12
US-08-961-083-66
Sequence 66, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 763 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-66

```

Query Match 63.5%; Score 33; DB 3; Length 763;
 Best Local Similarity 62.5%; Pred. No. 1.3e+02;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 3 VVPXGXHY 10
 DB 184 IVPXGDHY 191

DB 184 IVPXGDHY 191

```

RESULT 13
US-09-536-784-66
Sequence 66, Application US/09536784
Patent No. 6573082
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 763 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-09-536-784-66

```

Query Match 63.5%; Score 33; DB 4; Length 763;
 Best Local Similarity 62.5%; Pred. No. 1.3e+02;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 3 VVPXGXHY 10
 DB 184 IVPXGDHY 191

```

RESULT 14
US-08-961-083-56
Sequence 56, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

```

COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 796 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-56

Query Match 63.5%; Score 33; DB 3; Length 796;
Best Local Similarity 62.5%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
: || ||
Db 185 IVPXGDHY 192

RESULT 15
US-09-536-784-56
Sequence 56, Application US/09536784
Patent No. 6573082
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 796 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-536-784-56

Query Match 63.5%; Score 33; DB 4; Length 796;
Best Local Similarity 62.5%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
: || ||
Db 185 IVPXGDHY 192

Search completed: December 22, 2003, 16:43:45
Job time : 10.1333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 16:40:14 ; Search time 20.8667 Seconds
(Without alignments)
98.451 Million cell updates/sec

Title: US-09-909-164-44
Perfect score: 52
Sequence: 1 EREVFXGXHY 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues
Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/prodata/2/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/prodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/prodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/prodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/prodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/prodata/2/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/prodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/prodata/2/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/prodata/2/pubppaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/prodata/2/pubppaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/prodata/2/pubppaa/US09_NEW_PUB.pep:*
13: /cgn2_6/prodata/2/pubppaa/US10_PUBCOMB.pep:*
14: /cgn2_6/prodata/2/pubppaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/prodata/2/pubppaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/prodata/2/pubppaa/US10C_NEW_PUB.pep:*
17: /cgn2_6/prodata/2/pubppaa/US60_NEW_PUB.pep:*
18: /cgn2_6/prodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	65.4	381	12	US-09-769-736-24
2	34	65.4	715	11	US-09-252-088-16
3	34	65.4	715	12	US-10-340-792-16
4	34	65.4	793	11	US-09-252-088-15
5	34	65.4	793	12	US-10-340-792-15
6	34	65.4	822	12	US-09-769-736-18
7	33	63.5	447	9	US-09-765-272-182
8	33	63.5	484	11	US-09-769-787-38
9	33	63.5	484	12	US-09-769-744-24
10	33	63.5	763	9	US-09-765-272-66
11	33	63.5	796	9	US-09-765-272-56
12	33	63.5	826	11	US-09-769-787-194
13	33	63.5	838	11	US-09-884-465A-8
14	33	63.5	840	11	US-09-884-465A-7
15	33	63.5	840	11	US-09-884-465A-10

16	33	63.5	1039	11	US-09-884-465A-6	Sequence 6, Appl1
17	32	61.5	153	12	US-10-029-386-32486	Sequence 32486, A
18	32	61.5	345	12	US-10-038-307-24	Sequence 24, Appl1
19	32	61.5	345	15	US-10-201-292-24	Sequence 24, Appl1
20	32	61.5	499	15	US-10-156-761-11332	Sequence 11332, A
21	32	61.5	519	11	US-09-934-455-164	Sequence 164, Appl
22	32	61.5	519	12	US-10-302-267-36	Sequence 36, Appl1
23	32	61.5	717	12	US-10-171-404A-38	Sequence 38, Appl1
24	32	61.5	728	12	US-10-171-404A-36	Sequence 36, Appl1
25	32	61.5	1463	15	US-10-176-847-22	Sequence 22, Appl1
26	32	61.5	1463	15	US-10-205-823-343	Sequence 343, App
27	31	59.6	44	10	US-09-881-752A-292	Sequence 292, App
28	31	59.6	217	15	US-10-156-761-13210	Sequence 13210, A
29	31	59.6	354	12	US-10-279-029-104	Sequence 104, App
30	31	59.6	354	12	US-10-219-810-31	Sequence 31, Appl1
31	31	59.6	364	15	US-10-177-229-8	Sequence 8, Appl1
32	31	59.6	387	9	US-09-815-242-10416	Sequence 10416, A
33	31	59.6	482	12	US-10-298-638-27	Sequence 27, Appl1
34	31	59.6	485	12	US-10-298-638-10	Sequence 10, Appl1
35	31	59.6	632	10	US-09-738-626-4919	Sequence 4919, Ap
36	31	59.6	653	11	US-09-820-843A-26	Sequence 26, Appl1
37	31	59.6	665	10	US-09-942-447-2	Sequence 2, Appl1
38	31	59.6	665	12	US-10-319-399-61	Sequence 61, Appl1
39	31	59.6	700	15	US-10-319-762-2	Sequence 2, Appl1
40	31	59.6	761	12	US-10-032-585-7521	Sequence 7521, Ap
41	31	59.6	1163	12	US-10-122-067-4	Sequence 4, Appl1
42	31	59.6	1485	12	US-10-262-794A-32	Sequence 32, Appl1
43	31	59.6	1485	15	US-10-242-056-32	Sequence 32, Appl1
44	31	59.6	1499	12	US-10-096-534-67	Sequence 67, Appl1
45	31	59.6	1499	12	US-10-122-067-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-09-769-736-24
Sequence 24, Application US/09769736
Publication No. US20030138775A1
GENERAL INFORMATION:
APPLICANT: Microbial Technics Limited
APPLICANT: Le Page, Richard WP
APPLICANT: Wells, Jeremy M
APPLICANT: Hamill, Sean B
TITLE OF INVENTION: Proteins
FILE REFERENCE: PWC/P21089WO
CURRENT APPLICATION NUMBER: US/09/769,736
CURRENT FILING DATE: 2003-02-14
PRIOR APPLICATION NUMBER: GB 9816335.5
PRIOR FILING DATE: 1998-07-27
PRIOR APPLICATION NUMBER: US 60/125163
PRIOR FILING DATE: 1999-03-19
NUMBER OF SEQ ID NOS: 212
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 24
LENGTH: 381
TYPE: PRT
ORGANISM: Streptococcus agalactiae
US-09-769-736-24
Query Match 65.4%; Score 34; DB 12; Length 381;
Best Local Similarity 75.0%; Pred. No. 98;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Cy 3 VVFXGXHY 10
Db 350 VVFXGXHY 357
RESULT 2
US-09-252-088-16
Sequence 16, Application US/09252088
Publication No. US20030031682A1

```

; GENERAL INFORMATION:
; APPLICANT: BRODEUR, Bernard R.
; APPLICANT: RIOUX, Clément
; APPLICANT: BOYER, Martine
; APPLICANT: CHARLEBOIS, Isabelle
; APPLICANT: HAMEL, Josée
; APPLICANT: MARTIN, Denis
; TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 8331-9002
; CURRENT APPLICATION NUMBER: US/09/252,088
; EARLIER FILING DATE: 1999-02-18
; EARLIER APPLICATION NUMBER: US/60/075,425
; EARLIER FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 715
; TYPE: PRT
; ORGANISM: group B streptococcus
US-09-252-088-16
```

```

Query Match          65.4%; Score 34; DB 11; Length 715;
Best Local Similarity 75.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      3 VVPXGXHY 10
      ||| |||
Db      243 VVPHGDHY 250
```

```

RESULT 3
US-10-340-792-16
; Sequence 16, Application US/10340792
; Publication No. US20030228323A1
; GENERAL INFORMATION:
; APPLICANT: BRODEUR, Bernard R.
; APPLICANT: BOYER, Martine
; APPLICANT: CHARLEBOIS, Isabelle
; APPLICANT: HAMEL, Josée
; APPLICANT: MARTIN, Denis
; TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 8331-9002
; CURRENT APPLICATION NUMBER: US/10/340,792
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: US/09/252,088B
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US/60/075,425
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 715
; TYPE: PRT
; ORGANISM: group B streptococcus
US-10-340-792-16
```

```

Query Match          65.4%; Score 34; DB 12; Length 715;
Best Local Similarity 75.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      3 VVPXGXHY 10
      ||| |||
Db      243 VVPHGDHY 250
```

```

RESULT 4
US-09-252-088-15
; Sequence 15, Application US/09252088
; Publication No. US20030031682A1
; GENERAL INFORMATION:
; APPLICANT: BRODEUR, Bernard R.
; APPLICANT: RIOUX, Clément
```

```

; APPLICANT: BOYER, Martine
; APPLICANT: CHARLEBOIS, Isabelle
; APPLICANT: HAMEL, Josée
; APPLICANT: MARTIN, Denis
; TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 8331-9002
; CURRENT APPLICATION NUMBER: US/09/252,088
; CURRENT FILING DATE: 1999-02-18
; EARLIER APPLICATION NUMBER: US/60/075,425
; EARLIER FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 793
; TYPE: PRT
; ORGANISM: group B streptococcus
US-09-252-088-15
```

```

Query Match          65.4%; Score 34; DB 11; Length 793;
Best Local Similarity 75.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      3 VVPXGXHY 10
      ||| |||
Db      321 VVPHGDHY 328
```

```

RESULT 5
US-10-340-792-15
; Sequence 15, Application US/10340792
; Publication No. US20030228323A1
; GENERAL INFORMATION:
; APPLICANT: BRODEUR, Bernard R.
; APPLICANT: RIOUX, Clément
; APPLICANT: BOYER, Martine
; APPLICANT: CHARLEBOIS, Isabelle
; APPLICANT: HAMEL, Josée
; APPLICANT: MARTIN, Denis
; TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 8331-9002
; CURRENT APPLICATION NUMBER: US/10/340,792
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: US/09/252,088B
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US/60/075,425
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 793
; TYPE: PRT
; ORGANISM: group B streptococcus
US-10-340-792-15
```

```

Query Match          65.4%; Score 34; DB 12; Length 793;
Best Local Similarity 75.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      3 VVPXGXHY 10
      ||| |||
Db      321 VVPHGDHY 328
```

```

RESULT 6
US-09-769-736-18
; Sequence 18, Application US/09769736
; Publication No. US20030138775A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hanniffy, Sean B
; TITLE OF INVENTION: Proteins
```

FILE REFERENCE: PWC/P21089WO
CURRENT APPLICATION NUMBER: US/09/769,736
CURRENT FILING DATE: 2003-02-14
PRIORITY APPLICATION NUMBER: GB 9816335.5
PRIORITY FILING DATE: 1998-07-27
PRIORITY APPLICATION NUMBER: US 60/125163
PRIORITY FILING DATE: 1999-03-19
NUMBER OF SEQ ID NOS: 212
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 18
LENGTH: 822
TYPE: PRT
ORGANISM: Streptococcus agalactiae
US-09-769-736-18

Query Match 65.4%; Score 34; DB 12; Length 822;
Best Local Similarity 75.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
|||
Db 350 VVPHGDHY 357

RESULT 7
US-09-765-272-182
Sequence 182, Application US/09765272
Patent No. US20020061545A1
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 182:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 amino acids
TYPE: amino acid
STRANDDRESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 182:
US-09-765-272-182

Query Match 63.5%; Score 33; DB 9; Length 447;
Best Local Similarity 62.5%; Pred. No. 1.8e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 3 VVPXGXHY 10

Db :|||
178 IVPHGHY 185

RESULT 8
US-09-769-787-38
Sequence 38, Application US/09769787
Publication No. US20030091577A1
GENERAL INFORMATION:
APPLICANT: Microbial Techniques Limited
APPLICANT: Gilbert, Christophe FG
APPLICANT: Hansbro, Philip M
TITLE OF INVENTION: Proteins
FILE REFERENCE: PWC/P21129WO
CURRENT APPLICATION NUMBER: US/09/769,787
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: GB 9816337.1
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: US 60/125164
PRIOR FILING DATE: 1999-03-19
NUMBER OF SEQ ID NOS: 388
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 38
LENGTH: 484
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-769-787-38

Query Match 63.5%; Score 33; DB 11; Length 484;
Best Local Similarity 62.5%; Pred. No. 2e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
|||
Db 202 IVPHGHY 209

RESULT 9
US-09-769-744A-24
Sequence 24, Application US/09769744A
Publication No. US20030134407A1
GENERAL INFORMATION:
APPLICANT: Le Page, Richard WF
APPLICANT: Wells, Jeremy M
APPLICANT: Hanniffy, Sean B
APPLICANT: Hansbro, Philip M
TITLE OF INVENTION: Proteins
FILE REFERENCE: PWC/P21122WO
CURRENT APPLICATION NUMBER: US/09/769,744A
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: PCT/GB99/02452
PRIOR FILING DATE: 1999-07-27
PRIOR APPLICATION NUMBER: GB 9816336.3
PRIOR FILING DATE: 1998-07-27
PRIOR APPLICATION NUMBER: US 60/125329
PRIOR FILING DATE: 1999-03-19
NUMBER OF SEQ ID NOS: 196
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 24
LENGTH: 484
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-769-744A-24

Query Match 63.5%; Score 33; DB 12; Length 484;
Best Local Similarity 62.5%; Pred. No. 2e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
|||
Db 202 IVPHGHY 209

RESULT 10
US-09-765-272-66
Sequence 66, Application US/09765272
Patent No. US20020061545A1
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 763 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-09-765-272-66
Query Match 63.5%; Score 33; DB 9; Length 763;
Best Local Similarity 62.5%; Pred. No. 3.2e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 3 VVPXGXY 10
DB 184 IVPFGDHY 191
RESULT 11
US-09-765-272-56
Sequence 56, Application US/09765272
Patent No. US20020061545A1
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 796 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-765-272-56
Query Match 63.5%; Score 33; DB 9; Length 796;
Best Local Similarity 62.5%; Pred. No. 3.4e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 3 VVPXGXY 10
DB 185 IVPFGDHY 192
RESULT 12
US-09-769-787-194
Sequence 194, Application US/09769787
Patent No. US20030091577A1
GENERAL INFORMATION:
APPLICANT: Microbial Technics Limited
APPLICANT: Hansbro, Christophe FG
TITLE OF INVENTION: Proteins
FILE REFERENCE: PWC/P21129NO
CURRENT APPLICATION NUMBER: US/09/769,787
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: GB 9816337.1
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: US 60/125164
PRIOR FILING DATE: 1999-03-19
NUMBER OF SEQ ID NOS: 368
SOFTWARE: Patentm Ver. 2.1
SEQ ID NO 194
LENGTH: 826
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-769-787-194
Query Match 63.5%; Score 33; DB 11; Length 826;
Best Local Similarity 62.5%; Pred. No. 3.5e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 3 VVPXGXY 10
DB 204 IVPFGDHY 211
RESULT 13
US-09-884-465A-8
Sequence 8, Application US/09884465A
Patent No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee

```

; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-8

```

```

Query Match      63.5%; Score 33; DB 11; Length 838;
Best Local Similarity 62.5%; Pred. No. 3.6e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      3 VVPXGXHY 10
      :|||
Db      203 IVPXGDHY 210

```

```

RESULT 14
US-09-884-465A-7
; Sequence 7, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 840
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-7

```

```

Query Match      63.5%; Score 33; DB 11; Length 840;
Best Local Similarity 62.5%; Pred. No. 3.6e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      3 VVPXGXHY 10
      :|||
Db      205 IVPXGDHY 212

```

```

RESULT 15
US-09-884-465A-10
; Sequence 10, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens

```

```

; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 840
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-10

```

```

Query Match      63.5%; Score 33; DB 11; Length 840;
Best Local Similarity 62.5%; Pred. No. 3.6e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      3 VVPXGXHY 10
      :|||
Db      182 IVPXGDHY 189

```

```

Search completed: December 22, 2003, 17:32:45
Job time : 21.9333 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 17:24:36 ; Search time 9.06667 Seconds

(without alignments)
116.675 Million cell updates/sec

Title: US-09-909-164-44

Perfect score: 52

Sequence: 1 BEVVPKXGRHYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR 76:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	67.3	233	2	T02590
2	35	67.3	308	2	A72207
3	35	67.3	360	2	E69086
4	35	67.3	1028	2	AP3286
5	35	67.3	1396	2	S36851
6	34	65.4	102	2	A42452
7	34	65.4	126	2	A25905
8	34	65.4	279	2	C75538
9	34	65.4	495	2	T28717
10	34	65.4	534	2	A69284
11	34	65.4	822	2	T46758
12	34	65.4	1057	2	F89892
13	33	63.5	46	2	VKLJST
14	33	63.5	124	1	VKLJST
15	33	63.5	156	2	D82618
16	33	63.5	252	2	AE2001
17	33	63.5	412	2	A48702
18	33	63.5	460	2	S69046
19	33	63.5	510	2	G86430
20	33	63.5	743	2	G38143
21	33	63.5	802	2	G95136
22	33	63.5	819	2	B95136
23	33	63.5	828	2	B98004
24	33	63.5	839	2	G95115
25	33	63.5	853	2	C97985
26	33	63.5	855	2	D98004
27	33	63.5	1039	2	H95115
28	33	63.5	1039	2	D97985
29	32	61.5	162	2	A70939

30	32	61.5	225	2	S57810	hypothetical prote
31	32	61.5	286	2	G66169	prohibitin 2 (lipo
32	32	61.5	311	2	H69194	GMP synthetase, su
33	32	61.5	322	2	AF3211	hypothetical prote
34	32	61.5	364	2	S47540	fructose-bisphosph
35	32	61.5	369	2	T17267	hypothetical prote
36	32	61.5	374	2	G69119	probable aspartate
37	32	61.5	425	2	T24111	hypothetical prote
38	32	61.5	454	2	G96568	probable non-photo
39	32	61.5	519	2	G84598	probable bZIP tran
40	32	61.5	536	2	A13544	aldehyde dehydroge
41	32	61.5	627	2	A69663	DNA mismatch repai
42	32	61.5	696	2	A91247	phage transposase
43	32	61.5	701	2	S61239	hypothetical prote
44	32	61.5	716	1	UC5061	macrophage-stimula
45	32	61.5	738	2	A48246	ethylene-response

ALIGNMENTS

RESULT 1

T02590
DNA binding protein ERBP-2 - common tobacco

C/Species: Nicotiana tabacum (common tobacco)

C/Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 21-Jul-2000

C/Accession: T02590

R/Name-Takagi, M.; Shinohe, H.

Plant Cell 7, 173-182, 1995

A/Title: Ethylene-inducible DNA binding proteins that interact with an ethylene respon

A/Reference number: Z14671; MUID:9527645; PMID:7756828

A/Accession: T02590

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-233 <OHM>

A/Cross-references: EMBL:D38126; NID:G790362; PIDN:BA07324.1; PID:G1208498

A/Experimental source: strain BY4; tissue-type leaf

Query Match 67.3%; Score 35; DB 2; Length 233;
Best Local Similarity 60.0%; Pred. No. 9.8;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPKXGRHYS 10
DB 90 QAVVPKGRHY 99

RESULT 2

A72207
fish proteinase activity modulator HfIK - Thermotoga maritima (strain MSB8)

C/Species: Thermotoga maritima

C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C/Accession: A72207

R/Nelson, K.B.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic

Garratt, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, W.S.; Phillips, C.A.; Richardson, T

C.M.

Nature 399, 323-329, 1999

A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A/Reference number: A72200; MUID:99287316; PMID:10360571

A/Accession: A72207

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-308 <ARN>

A/Cross-references: GB:AB001819; GB:AB000512; NID:G4982396; PIDN:AD36885.1; PID:G4982

A/Experimental source: strain MSB8

C/Genetics:

A/Genes: TM1822

C/Superfamily: erythrocyte band 7 integral membrane protein

Query Match 67.3%; Score 35; DB 2; Length 308;
Best Local Similarity 75.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 VVVPXGXHY 10
 DB 41 VVPSGIMY 48

RESULT 3

cell division protein - Methanobacterium thermoautotrophicum (strain Delta H)
 C/Species: Methanobacterium thermoautotrophicum
 C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
 A/Accession: B69086
 R/Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
 ; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
 Ki, S.; Church, G.M.; Daniels, C.V.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A/Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: func
 A/Reference number: A69000; MUID:98037514; PMID:9371463
 A/Accession: B69086
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-360 <MTH>
 A/Cross-references: GB:AE000923; GB:AE000666; NID:92622766; PIDN:AMB86115.1; PID:9262277
 A/Experimental source: strain Delta H
 C/Genetics:
 A/Gene: MTH1642
 C/Superfamily: cell division protein MTH174

Query Match 67.3%; Score 35; DB 2; Length 360;
 Best Local Similarity 45.5%; Pred. No. 16;
 Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 EEVVPXGXHY 11
 DB 98 EDLVPMSHHT 108

RESULT 4

ATP-dependent DNA helicase BMEI0275 [imported] - Brucella melitensis (strain 16M)
 C/Species: Brucella melitensis
 C/Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
 A/Accession: AF3286
 R/Dalvechio, V.G.; Kaprat, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova,
 ; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letes
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A/Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
 A/Reference number: AD3252; PMID:11756688
 A/Accession: AF3286
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-1028 <KUR>
 A/Cross-references: GB:AE006917; PIDN:AAU51457.1; PID:917982167; GSPDB:GN00190
 A/Experimental source: strain 16M
 C/Genetics:
 A/Gene: BMEI0275
 A/Map position: 1

Query Match 67.3%; Score 35; DB 2; Length 1028;
 Best Local Similarity 54.5%; Pred. No. 49;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 EEVVPXGXHY 11
 DB 76 EKIIVPGARYS 86

RESULT 5

536851
 L-shaped tail fiber protein - phage T5
 N/Alternate names: ltf protein
 C/Species: phage T5
 C/Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 17-Nov-2000
 C/Accession: S65934; S01984; S36851

R/Kallman, A.V.; Kulshin, V.E.; Shlyapnikov, M.G.; Kezenko, V.N.; Kryukov, V.M.
 FEBS Lett. 366, 46-48, 1995
 A/Title: The nucleotide sequence of the bacteriophage T5 ltf gene.
 A/Reference number: S65934; MUID:95309401; PMID:7789514
 A/Accession: S65934
 A/Molecule type: DNA
 A/Residues: 1-1396 <KAL>

A/Cross-references: EMBL:X69460; NID:915415; PIDN:CAA49220.1; PID:915416
 R/Kallman, A.V.; Kryukov, V.M.; Bayev, A.A.
 Nucleic Acids Res. 16, 6230, 1988

A/Title: The nucleotide sequence of bacteriophage T5 DNA at the region between early a
 A/Reference number: S01982; MUID:88289370; PMID:3467228
 A/Accession: S01984

A/Status: translation not shown

A/Molecule type: DNA

A/Residues: 934-985, 'A', 987-1396 <KA2>

A/Cross-references: EMBL:X07559

C/Genetics:

A/Gene: ltf

C/Keywords: late protein; tail fiber

Query Match 67.3%; Score 35; DB 2; Length 1396;
 Best Local Similarity 50.0%; Pred. No. 68;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 EYVVPXGXHY 11
 DB 1360 KTIIPAGDHY 1369

RESULT 6

VI protein - tobacco yellow dwarf virus (strain Australia)
 C/Species: tobacco yellow dwarf virus
 C/Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
 A/Accession: A42452
 R/Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.B.
 Virology 187, 633-642, 1992
 A/Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yell
 A/Reference number: A42452; MUID:92188538; PMID:1546458
 A/Accession: A42452
 A/Molecule type: DNA
 A/Residues: 1-102 <MOR>
 A/Cross-references: GB:M81103; NID:9335283; PIDN:AAA47947.1; PID:9335284

Query Match 65.4%; Score 34; DB 2; Length 102;
 Best Local Similarity 60.0%; Pred. No. 6.5;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 EYVVPXGXHY 11
 DB 7 QVVPSSGIMYS 16

RESULT 7

A25905
 tachykinin B precursor - bovine
 N/Alternate names: neuromedin K
 C/Species: Bos primigenius taurus (cattle)
 C/Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 16-Jul-1999
 C/Accession: A25905
 R/Kotani, H.; Hoshimaru, M.; Nawa, H.; Nakanishi, S.
 Proc. Natl. Acad. Sci. U.S.A. 83, 7074-7078, 1986
 A/Title: Structure and gene organization of bovine neuromedin K precursor.
 A/Reference number: A25905; MUID:86313713; PMID:3462746
 A/Accession: A25905
 A/Molecule type: DNA
 A/Residues: 1-126 <KOT>
 A/Cross-references: GB:M14351; NID:9163587; PIDN:AAA30723.1; PID:9163590
 C/Superfamily: neurokinin B precursor
 F/1-26/Domain: signal sequence #status predicted <SIG>
 F/27-126/Product: proteachykinin B #status predicted <WAT>

Query Match 65.4%; Score 34; DB 2; Length 126;
 Best Local Similarity 66.7%; Pred. No. 8.2;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGXKH 9
 |||||
 DB 28 EQVPRGGGH 36

RESULT 8

hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000

C:Accession: C75538

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: C75538

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-279 <WHI>

A:Cross-references: GB:AB001889; GB:AB000513; NID:96457944; PIDN:AA09867.1; PID:9645795

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0271

A:Map position: 1

C:Superfamily: Deinococcus radiodurans hypothetical protein DR0271

Query Match 65.4%; Score 34; DB 2; Length 279;
 Best Local Similarity 75.0%; Pred. No. 19;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 VPKXGKHS 11
 |||||
 DB 100 VPLGRHYS 107

RESULT 9

hypothetical protein F10D2.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C:Accession: T28717

R:Graves, T.; Wohldmann, P.; Gillam, B.

submitted to the EMBL Data Library, September 1997

A:Description: The sequence of C. elegans cosmid F10D2.

A:Reference number: Z20515

A:Accession: T28717

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-495 <GRA>

A:Cross-references: EMBL:AF022972; PIDN:AA048234.1; GSPDB:GN00023; CESP:F10D2.3

A:Experimental source: strain Bristol N2; clone F10D2

C:Genetics:

A:Gene: CESP:F10D2.3

A:Map position: 5

A:introns: 37/2; 90/1; 113/1; 183/3; 356/3; 380/3; 428/2

Query Match 65.4%; Score 34; DB 2; Length 495;
 Best Local Similarity 50.0%; Pred. No. 36;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXKH 10
 |||||
 DB 218 ENIVPTGKH 227

RESULT 10

A69284

coenzyme F420-guione oxidoreductase (BC 1.6.5.-) 56k chain - Archaeoglobus fulgidus

N:Alternate names: sarcosine oxidase alpha chain cox4 homolog

C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 19-May-2000

C:Accession: A69284; S45665

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.B.; Ketchum, K.A.; Doi,

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes,

Smith, H.O.; Moese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch.

A:Reference number: A69284; MUID:98049343; PMID:9389475

A:Accession: A69284

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-534 <KLB>

A:Cross-references: GB:AB001086; GB:AB000782; NID:92689409; PIDN:AA090959.1; PID:9265

R:Kunow, J.; Linder, D.; Stetter, K.O.; Thauer, R.K.

EUR. J. Biochem. 223, 503-511, 1994

A:Title: F(420)H(2): guinone oxidoreductase from Archaeoglobus fulgidus. Characteriza

A:Reference number: S45665; MUID:94333340; PMID:8055920

A:Accession: S45665

A:Molecule type: Protein

A:Residues: 2,'X','4','X','6-7','X','9-10','XX','13-14 <KUN>

A>Note: the authors could not distinguish between glutamate and cysteine for residues

C:Superfamily: glutamate synthase small chain

C:Keywords: oxidoreductase

Query Match 65.4%; Score 34; DB 2; Length 534;
 Best Local Similarity 45.5%; Pred. No. 39;
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXKH 11
 |||||
 DB 119 DKVPRGSHYT 129

RESULT 11

hypothetical 92.4k protein - Streptococcus agalactiae

C:Species: Streptococcus agalactiae

C:Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000

C:Accession: T46758

R:Spillierberg, B.; Rozdzinski, B.; Martin, S.; Weber-Heymann, J.; Schnitzler, N.; La

Infect. Immun. 67, 871-878, 1999

A:Title: Imb, a protein with similarities to the Irai adhesin family, mediates attachm

A:Reference number: Z24091; MUID:99115568; PMID:9916102

A:Accession: T46758

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-822 <SPE>

A:Cross-references: EMBL:AF062533; NID:94249622; PIDN:AA013797.1; PID:94249624

A:Experimental source: strain R268

C:Superfamily: Streptococcus agalactiae hypothetical 92.4k protein

QY 3 VVPXGXKH 10
 |||||
 DB 350 VVPXGXKH 357

Query Match 65.4%; Score 34; DB 2; Length 822;
 Best Local Similarity 75.0%; Pred. No. 62;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 12
 P89892
 cardamoyl-phosphate synthase large chain [imported] - Staphylococcus aureus (strain N:
 C:Species: Staphylococcus aureus
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C:Accession: P89892
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yurawa, H.; Kobayashi, I.; Cui, L.; Oe
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.

C.: Shiba, T.; Hattori, M.; Ogasawara, N.; Hayaishi, H.; Hiramatsu, K.
 Lancet 357, 1225-1240, 2001
 A>Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
 A:Reference number: A89758; MUID:21311952; PMID:11418146
 A:Accession: F89892
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1057 <KUR>
 A:Cross-references: GB:BA000018; PID:913701002; PIDN:BAB42298.1; GSPDB:GN00149
 A:Experimental source: strain N315
 C:Genetics:
 A:Gene: *pyrAB*
 C:Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain; biotin

Query Match 65.4%; Score 34; DB 2; Length 1057;
 Best Local Similarity 60.0%; Pred. No. 82;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGXHS 11
 |||
 Db 190 EIVSNGDHY 199

RESULT 13
 E97985
 hypothetical protein spr0909 [imported] - *Streptococcus pneumoniae* (strain R6)
 C:Species: *Streptococcus pneumoniae*
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
 C:Accession: E97985
 R:Hoeking, J.A.; Alborn, Jr., W.; Arnold, J.; Blaszcak, L.; Burgelt, S.; Dehoff, B.S.; E
 r, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
 y, P.; Sun, P.M.; Winler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Taskunas, S.R.;
 A>Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.
 A:Reference number: A97872; MUID:21429245; PMID:11544234
 A:Accession: E97985
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-46 <KUR>
 A:Cross-references: GB:AE007317; PIDN:AAK9713.1; PID:G15458516; GSPDB:GN00174
 C:Genetics:
 A:Gene: *spr0909*

Query Match 63.5%; Score 33; DB 2; Length 46;
 Best Local Similarity 62.5%; Pred. No. 4.5;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXH 10
 |||
 Db 21 IVPHGHY 28

RESULT 14
 VKLJ51
 trans-regulatory splicing protease - simian immunodeficiency virus STVcpz
 N:Alternate names: anti-repression trans-activator; art protein; rev protein; trs protei
 C:Species: simian immunodeficiency virus STVcpz
 A>Note: host Pan troglodytes (chimpanzee)
 C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
 C:Accession: S09988
 R:Huet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Main-Hobson, S.
 Nature 345, 356-359, 1990
 A>Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.
 A:Reference number: S09983; MUID:90259077; PMID:2188136
 A:Accession: S09988
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-124 <HUB>
 A:Cross-references: EMBL:X52154; NID:G58866; PIDN:CAA36405.1; PID:G763085
 C:Genetics:
 A:Gene: rev; trs; art
 A:Introns: 27/1

C:Superfamily: AIDS trans-regulatory splicing protein
 C:Keywords: AIDS; immunodeficiency; splicing protein; transcription regulation

Query Match 63.5%; Score 33; DB 1; Length 124;
 Best Local Similarity 60.0%; Pred. No. 13;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGXHS 11
 |||
 Db 107 EIVPAGGNY 116

RESULT 15
 D82618
 conserved hypothetical protein Xr1950 [imported] - *Xylella fastidiosa* (strain 945c)
 C:Species: *Xylella fastidiosa*
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: D82618
 R:Anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequ
 Nature 406, 151-157, 2000
 A>Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
 A:Reference number: A82515; MUID:20365717; PMID:10910347
 A>Note: for a complete list of authors see reference number A59328 below
 A:Accession: D82618
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-156 <SIM>
 A:Cross-references: GB:AE004014; GB:AE003849; NID:G9107044; PIDN:AAF84752.1; GSPDB:GNO
 A:Experimental source: strain 945c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvares, R.;
 Brites, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carter,
 de-Neto, E.; Docena, C.; El-Dorry, H.; Facinelli, A.P.; Ferreira, A.U.S.
 Submitted to Genbank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Pro
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.B.; Lai
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Savas
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silve
 M.; Tshakko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.;
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: Xr1950

Query Match 63.5%; Score 33; DB 2; Length 156;
 Best Local Similarity 55.6%; Pred. No. 17;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXH 9
 |||
 Db 119 ERIIPQGVH 127

Search completed: December 22, 2003, 17:45:00
 Job time : 9.0666 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:43:51 ; Search time 4.6 Seconds

(without alignments)
112.455 Million cell updates/sec

Title: US-09-909-164-44
Perfect score: 52
Sequence: 1 BEVVPXGXHY 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	69.2	426	AROQ_VIBPA	Q87GX9 vibrio para
2	35	67.3	1058	CARB_FUSNN	Q87G86 fusobacteri
3	35	67.3	1396	VLPF_BPTS	P13390 bacterioph
4	34	65.4	102	YLIK_TYDVA	P31619 tobacco yel
5	34	65.4	126	TKNK_BOVIN	P08858 bos taurus
6	34	65.4	1057	CARB_STAMW	Q95940 streptococ
7	34	65.4	1057	CARB_STAMW	P17280 chimpanzee
8	33	63.5	124	REV_STVCZ	Q03010 saccharomyc
9	33	63.5	460	IMEI_YEAST	Q89A96 buchera ap
10	33	63.5	578	MDLB_BUCBP	P36148 saccharomyc
11	33	63.5	743	YK47_YEAST	O26806 methanobact
12	32	61.5	308	GABF_METTH	P52210 ovis aries
13	32	61.5	363	ALFB_SHEEP	Q9WMP4 escherichia
14	32	61.5	573	SUOX_DROME	P48850 bacillus su
15	32	61.5	627	MTUL_BACUL	O4230 braesica ol
16	32	61.5	735	ETRI_BROUL	P43333 arabidopsis
17	32	61.5	738	ETRI_BROUL	P43333 arabidopsis
18	32	61.5	906	CENK_MOUSE	P49452 mus musculu
19	32	61.5	1374	YOS6_CABEL	Q05541 caenorhabdi
20	32	61.5	1378	RON_MOUSE	Q61190 mus musculu
21	31	59.6	130	SZOF_RAT	P97885 rattus norv
22	31	59.6	212	MSRA_VIBCH	Q9FDP0 vibrio chol
23	31	59.6	319	YHAI_CRYPA	P05062 homo sapien
24	31	59.6	363	ALFB_HUMAN	P79226 oryctolagus
25	31	59.6	363	ALFB_RABIT	Q84814 escherichia
26	31	59.6	387	THIK_BCO57	P21151 escherichia
27	31	59.6	387	THIK_BCO57	Q91616 salmonella
28	31	59.6	387	THIK_SALTY	Q91616 salmonella
29	31	59.6	426	SUSI_YARLI	Q91616 salmonella
30	31	59.6	427	TOLB_HAELIN	P46231 vibrio para
31	31	59.6	441	YL15_VIBPA	P40848 schistosacch
32	31	59.6	991	DHP1_SCHPO	P34335 caenorhabdi
33	31	59.6	1188	KPBA_CABEL	P34335 caenorhabdi

34	31	59.6	1410	1	PDRB_YEAST	P40550 saccharomyc
35	31	59.6	1499	1	ALOC_HUMAN	Q60312 homo sapien
36	31	59.6	3099	1	POLG_PRRYM	O56075 p genome po
37	30.5	58.7	268	1	ILIA_BOVIN	P08831 bos taurus
38	30.5	58.7	268	1	ILIA_CAPII	P79161 capra hircu
39	30.5	58.7	268	1	ILIA_SHEEP	Q28579 ovis aries
40	30	57.7	121	1	TKNK_HUMAN	Q9H10 homo sapien
41	30	57.7	152	1	Y16D_BPTA	P22917 bacterioph
42	30	57.7	160	1	Y201_MERTU	Q27654 methanococ
43	30	57.7	223	1	COAT_CIV36	Q00686 citrus tris
44	30	57.7	232	1	SCOA_HELPY	Q92163 helicobacte
45	30	57.7	232	1	SCOA_HELPY	P56006 helicobacte

ALIGNMENTS

RESULT 1

AROQ_VIBPA STANDARD; PRT; 426 AA.

AC Q87GX9; 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 3-phosphohistidine-1-carboxyvinyltransferase (EC 2.5.1.19) (5-enolpyruvylethylamine-3-phosphate synthase) (BSP synthase) (BSPS).
 GN AROQ OR VP1020.
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio.
 CC NCBI_TaxID=670;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIMD 2210633 / Serotype O3:K6;
 RX MEDLINE=22508454; PubMed=12620739;
 RA Martino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K., Iijima Y., Naito M., Nakano M., Yamashita A., Kubota Y., Kimura S., Yanagawa T., Honda T., Shinagawa H., Hattori M., Iida T.;
 RA Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism distinct from that of V. cholerae.
 RT Lancet 361:743-749(2003).
 RL -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphohistidine = phosphate + 5-O-(1-carboxyvinyl)-3-phosphohistidine.
 CC -1- PATHWAY: Aromatic amino acids biosynthesis; histidine pathway; sixth step.
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- SIMILARITY: Belongs to the BSP synthase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).
 CC EMBL; AP005076; BAC59283.1; -
 DR HAMAP; MF 00210; -; 1.
 DR PROSITE; PS00104; BSP SYNTHASE 1; 1.
 DR PROSITE; PS00885; BSP SYNTHASE 2; 1.
 KW Aromatic amino acid biosynthesis; Transferase; Complete proteome.
 SQ SEQUENCE 426 AA; 46094 MW; 373039C5B1A1F0F CRC64;

Query Match 69.2%; Score 36; DB 1; Length 426;
 Best Local Similarity 60.0%; Pred. No. 5.3;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXHY 10
 DB 223 BEVVPXGXHY 232

RESULT 2
CARB_FUSNN STANDARD; PRT; 1058 AA.
ID CARB_FUSNN
AC OBRG86;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (carbamoyl-phosphate synthetase ammonia chain).
GN CARB OR FN0422.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxId=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=2186394; PubMed=11889109;
RA Kapralov V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A., Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L., Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A., Larsen N., D'Souza M., Malunas T., Pusch G., Haselkorn R., Forslein M., Kyriades N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.
CC -1- COPACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -1- PATHWAY: Arginine biosynthesis.
CC -1- PATHWAY: Pyrimidine biosynthesis, first step.
CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).
CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AB010554; AAL94625.1; ALT_INIT.
DR HAMAP: MF_01210; -; 1
DR InterPro: IPR006273; CarA_L_glu.
DR InterPro: IPR005483; CPase_L.
DR InterPro: IPR005479; CPase_L_D2.
DR InterPro: IPR005480; CPase_L_D3.
DR InterPro: IPR005481; CPase_L_N.
DR InterPro: IPR004362; MGS_1Lk.
DR Pfam: PF00289; CPase_L_chain; 2.
DR Pfam: PF02786; CPase_L_D2; 2.
DR Pfam: PF02787; CPase_L_D3; 1.
DR Pfam: PF02142; MGS; 1.
DR PRINTS: PRO0098; CPASAB.
DR TIGRfam: TIGR01369; CPASAB1_lrg; 1.
DR PROSITE: PS00866; CPASAB_1; 2.
DR PROSITE: PS00867; CPASAB_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat; ATP-binding; Manganese; Complete proteome.
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1058
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).

FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1B39F CRC64;
Query Match 67.3%; Score 35; DB 1; Length 1058;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 2 EVVFXGHYS 11
Db 190 EIVPNGIAYS 199

RESULT 3
ID VLPF_BPTS STANDARD; PRT; 1396 AA.
AC P13350; O48502;
DT 01-JUN-1990 (Rel. 13, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE L-shaped tail fiber protein (LTF protein).
GN LTF.
OS Bacteriophage T5.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC T5-like viruses.
OX NCBI_TaxId=10726;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95309401; PubMed=1789514;
RA Kaliman A.V., Kuleshin V.B., Shlyapnikov M.G., Ksenzenko V.N., Kyrkov V.M.;
RT "The nucleotide sequence of the bacteriophage T5 ltf gene."
RL FEBS Lett. 366:46-48(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Kaliman A.V.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=88289370; PubMed=3267228;
RA Kaliman A.V., Kyrkov V.M., Bayev A.A.;
RT "The nucleotide sequence of bacteriophage T5 DNA at the region between early and late genes."
RL Nucleic Acids Res. 16:6230-6230(1988).
CC -1- FUNCTION: NONESSENTIAL PROTEIN THAT MEDIATES BINDING TO THE POLYMANNOSE O ANTIGEN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: X69460; CA49220.1; -
DR EMBL: AJ001191; CA404591.1; -
DR PIR: S65934; S36851.
KW Late protein.
FT CONFLICT 986 986 V -> A (IN RRF, 2).
SQ SEQUENCE 1396 AA; 147989 MW; 18CD2192B5FFFC1 CRC64;
Query Match 67.3%; Score 35; DB 1; Length 1396;
Best Local Similarity 50.0%; Pred. No. 29;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 2 EVVFXGHYS 11
Db 1360 KTIAPGDHYS 1369

RESULT 4

```

Y11K TYDVA
ID Y11K TYDVA STANDARD; PRT; 102 AA.
AC P31619;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Hypothetical 11.2 kDa protein.
GN V1.
OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
OX NCBI_TaxID=31599;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92188538; PubMed=1546458;
RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.B.;
RT "The nucleotide sequence of the infectious cloned DNA component of
RT tobacco yellow dwarf virus reveals features of geminiviruses
RT infecting monocotyledonous plants."
RT Virology 187:633-642(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M81103; AAA47947.1; -.
DR PIR; A42452; A42452.
DR InterPro; IPR002621; Gemin1 mov.
DR Pfam; PF01708; Gemin1_mov; 1.
DR Hypothetical protein.
SQ SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF5B67 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 102;
Best Local Similarity 60.0%; Pred. No. 3.2;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPGXGHS 11
   ||| |
   7 QVVPGGINYS 16

RESULT 5
TRNK_BOVIN
ID TRNK_BOVIN STANDARD; PRT; 126 AA.
AC P08858;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurokinin B precursor (NBK) (Neuromedin K).
GN TAC3 OR NBKB.
OS Bos taurus (bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86313713; PubMed=3462746;
RA Koranyi H., Hoshimaru M., Nawa H., Nakanishi S.;
RT "Structure and gene organization of bovine neuromedin K precursor.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:7074-7078(1986).
CC -1- FUNCTION: TACHYKININS ARE ACTIVE PEPTIDES WHICH EXCITE NEURONS,
CC EVOKE BEHAVIORAL RESPONSES, ARE POTENT VASODILATORS AND
CC SECRETAGOGUES, AND CONTRACT (DIRECTLY OR INDIRECTLY) MANY SMOOTH
CC MUSCLES.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE TACHYKININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M14351; AAA30723.1; -.
DR EMBL; M14347; AAA30723.1; JOINED.
DR EMBL; M14349; AAA30723.1; JOINED.
DR EMBL; M14349; AAA30723.1; JOINED.
DR EMBL; M14350; AAA30723.1; JOINED.
DR PIR; A25905; A25905.
DR InterPro; IPR003635; Neurokinin.
DR InterPro; IPR002040; Tachykinin.
DR Pfam; PF03823; Neurokinin_B; 1.
DR ProDom; PD020370; Neurokinin; 1.
DR PROSITE; PS00267; TACHYKININ; 1.
KW Tachykinin; Neuropeptide; Cleavage on pair of basic residues;
KW Amidation; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT PROPE 21 83
FT PREPIDE 86 95 NEUROKININ B.
FT PROPE 99 126
FT MOD_RES 95 95
SQ SEQUENCE 126 AA; 13871 MW; 446BP433498C059 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 126;
Best Local Similarity 66.7%; Pred. No. 4;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXH 9
   ||| |
   28 EQVPGGCH 36

RESULT 6
CARB_STAM
ID CARB_STAM STANDARD; PRT; 1057 AA.
AC Q99UR5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (BC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
GN CARB OR PYRAB OR SAV1203 OR SA1046.
OS Staphylococcus aureus (strain M50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=M50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
RA Kanemori M., Matsunaru H., Maruyama A., Murakami H., Hoshoyama A.,
RA Sekimizu K., Hirakawa H., Kunara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yashashita A., Oshima K., Furuya K., Kaito C.,
RA Hattori M., Ogasawara N., Hayaishi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -1- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -1- PATHWAY: Arginine biosynthesis.
CC -1- PATHWAY: Pyrimidine biosynthesis, first step.
CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -----
CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP003361; BAB57365.1; -
DR EMBL; AP003312; BAB42298.1; -
DR PIR; P89892; P89892.
DR HSBP; P00968; ICS0.
DR HAMAP; MF_01210; -; 1.
DR InterPro; IPR006275; CARA_L_glu.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR004362; MGS_1like.
DR Pfam; PF00289; CPase_L_chain; 2.
DR Pfam; PF02786; CPase_L_D2; 2.
DR Pfam; PF02787; CPase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPASAB.
DR TIGRfam; TIGR01369; CPaseII_1rg; 1.
DR PROSITE; PS00866; CPASAB_1; 2.
DR PROSITE; PS00867; CPASAB_2; 2.
DR Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
DR ATP-binding; Manganese; Complete proteome.
DR DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1057 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1057
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1057 AA; 117171 MW; E3E179EF0591F0F8 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 1057;
Best Local Similarity 60.0%; Pred. No. 36;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPGXGKHS 11
|:|:|:|:|
Db 190 EIVSNGLHYS 199

RESULT 7
CARB_STRAW STANDARD; PRT; 1057 AA.
AC P58940;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
GN CARB OR PYRAB OR MW1086.
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
ON NCB1_TaxID=196620;
RX MEDLINE=22040717; PubMed=12044378;
RA Babu T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwana K., Asano K., Naini T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.

```

```

RT "genome and virulence determinants of high virulence community-
RT acquired MRSA."
RL Lancet 359:1819-1827(2002).
CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -1- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -1- PATHWAY: Arginine biosynthesis.
CC -1- PATHWAY: Pyrimidine biosynthesis, first step.
CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP004825; BAB94951.1; -
DR HAMAP; MF_01210; -; 1.
DR InterPro; IPR006275; CARA_L_glu.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR004362; MGS_1like.
DR Pfam; PF00289; CPase_L_chain; 2.
DR Pfam; PF02786; CPase_L_D2; 2.
DR Pfam; PF02787; CPase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPASAB.
DR TIGRfam; TIGR01369; CPaseII_1rg; 1.
DR PROSITE; PS00866; CPASAB_1; 2.
DR PROSITE; PS00867; CPASAB_2; 2.
DR Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
DR ATP-binding; Manganese; Complete proteome.
DR DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1057 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1057
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1057 AA; 117185 MW; DBE3B09F9BC6F152 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 1057;
Best Local Similarity 60.0%; Pred. No. 36;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPGXGKHS 11
|:|:|:|:|
Db 190 EIVSNGLHYS 199

RESULT 8
REV_STVVCZ STANDARD; PRT; 124 AA.
AC P17280;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE REV protein (Anti-repression transactivator protein) (ART/TRS).
GN REV.

```

OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
 OC Viruses; Retroviral viruses; Retroviridae; Lentiviruses.
 OX NCBI_TaxID=11723;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90259077; Pubmed=2188136;
 RA Huot T., Cheyrier R., Meyerhans A., Roelants G., Wain-Hobson S.;
 RT "Genetic organization of a chimpanzee lentivirus related to HIV-1.";
 RL Nature 345:356-359(1990).
 CC -1- FUNCTION: REV APPEARS TO ACT POST-TRANSCRIPTIONALLY TO RELIEVE
 CC NEGATIVE REPRESSION OF GAG AND ENV PRODUCTION.
 CC -1- SUBCELLULAR LOCATION: Nuclear; accumulates in the nucleoli.
 CC -1- PM: PHOSPHOPROTEIN WHOSE STATE OF PHOSPHORYLATION IS MEDIATED
 CC BY A SPECIFIC SERINE KINASE ACTIVITY PRESENT IN THE NUCLEUS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X52154; CAA36405.1; -.
 CC DR EMBL: X52154; VKLJST.
 CC DR PIR: S09988; VKLJST.
 CC DR HIV: X52154; REVSCPZ.
 CC DR InterPro: IPR000625; REV_protein.
 CC Pfam: PF00424; REV; 1.
 CC KW Transcription regulation; AIDS; Phosphorylation; Nuclear protein.
 CC SQ SEQUENCE 124 AA; 13701 MW; F5877D1BDF65A7B2 CRC64;
 Query Match 63.5%; Score 33; DB 1; Length 124;
 Best Local Similarity 60.0%; Pred. No. 6.5;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 EVVPGXGHTS 11
 Db 107 ETVPAAGNYTS 116
 RESULT 9
 UME1_YEAST STANDARD; PRT; 460 AA.
 ID UME1_YEAST
 AC 003010; P87330;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Meiosis negative regulator UME1.
 GN UME1 OR WTM3 OR YPL139C OR LPI7C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=A364A;
 RA Mallory M.J., Strich R.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 CC [2]
 CC SEQUENCE FROM N.A.
 CC STRAIN=8288C / AB972;
 CC MEDLINE=97312721; Pubmed=9169875;
 CC Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,
 CC Azeiteiro J., Aparicio A., Bartell B.G., Badcock K., Benes V.J.,
 CC Bockstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
 CC Chung B., Churcher C.M., Coster P., Davis K., Davis R.W.,
 CC Dietrich F.S., Deltus H., DiPaolo T., Dubois E., Duesterhoeft A.,
 CC Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
 CC Hall J., Hedling U., Heumann K., Hilbert H., Hillier L.,
 CC Hannick-Smith S., Hyman R., Johnston M., Kaiman S., Kleene K.,
 CC Kamp C., Kurat O., Lashkari D., Lew H., Lin A., Lin D., Louis B.J.,
 CC Marathe C., Messenguy F., Mewes H.-W., Mitrapati S., Moestl D.,
 CC Mueller-Auer S., Namath A., Neuwirth U., Oefner P., Pearson D.,

RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
 RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tetzlaff H.,
 RA Trestarazu L.A., Uehinsky S., Vierendeels F., Vissers S., Voss H.,
 RA Walsh S.W., Wambutt R., Wang Y., Wedler E., Wedler H., Wilmott E.,
 RA Zhong W.W., Zolner A., Vo D.H., Hani J.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
 RL Nature 387:103-105(1997).
 CC -1- FUNCTION: TRANSCRIPTIONAL MODULATOR WITH ROLES IN MEIOTIC
 CC REGULATION AND SILENCING. NEGATIVE REGULATOR OF MEIOSIS.
 CC -1- SIMILARITY: Contains 4 WD repeats.
 CC -1- SIMILARITY: STRONG, TO YEAST WTM1 AND WTM2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U10280; AAB40937.1; -.
 CC DR EMBL: U43703; AAB68221.1; -.
 CC DR PIR: S69046; S69046.
 CC DR TRANSPAC: T04309; -.
 CC DR SGD: S0006060; UME1.
 CC DR GO: GO:0005634; C:nucleus; IDA.
 CC DR GO: GO:0003714; P:transcription co-repressor activity; IDA.
 CC DR GO: GO:0040020; P:regulation of meiosis; IGI.
 CC DR InterPro: IPR001680; WD40.
 CC Pfam: PF00400; WD40; 3.
 CC SMART: SM00320; WD40; 4.
 CC DR PROSITE: PS00678; WD_REPEATS_1; FALSE NEG.
 CC DR PROSITE: PS50082; WD_REPEATS_2; FALSE NEG.
 CC DR PROSITE: PS50294; WD_REPEATS_REGION; FALSE NEG.
 CC KW Transcription regulation; Meiosis; Repeat; WD repeat.
 CC FT REPEAT 233 271 WD 1.
 CC FT REPEAT 276 316 WD 2.
 CC FT REPEAT 339 379 WD 3.
 CC FT REPEAT 411 451 WD 4.
 CC SQ SEQUENCE 460 AA; 51022 MW; AA6F60448B7BCBA9 CRC64;
 Query Match 63.5%; Score 33; DB 1; Length 460;
 Best Local Similarity 62.5%; Pred. No. 25;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 3 VVPGXGHTY 10
 Db 85 IVPGLGHTY 92
 RESULT 10
 MDLB_BUCBP STANDARD; PRT; 578 AA.
 ID MDLB_BUCBP
 AC 089A96;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Multidrug resistance-like ATP-binding protein mdlb.
 GN MDLB OR BBP424.
 OS Buchnera aphidicola (subsp. Balgonga pteractiae).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Buchnera.
 OX NCBI_TaxID=135842;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2426901; Pubmed=12522265;
 RA Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F.,
 RA Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva P.J.,
 RA Tamames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.;
 RT "Reductive genome evolution in Buchnera aphidicola.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; A014017; A027134.1; -
DR PROSITE; PS00211; ABC_TRANSPORTER_1; FALSE_NEG.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KM ATP-binding; Transport; Transmembrane; Complete proteome.
FT TRANSMEM 26 46 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 166 186 POTENTIAL.
FT TRANSMEM 196 216 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT DOMAIN 339 573 ABC_TRANSPORTER.
FT NP_BIND 373 380 ATP (POTENTIAL).
SQ SEQUENCE 578 AA; 66827 MW; 3B84848CE196ADF7 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 578;
Best Local Similarity 50.0%; Pred. No. 31;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPGXHY 11
Db 550 RIVESGTHN 559

RESULT 11
YK47 YEAST STANDARD; PRT; 743 AA.
ID YK47 YEAST STANDARD; PRT; 743 AA.
AC P31148;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 83.6 kDa protein in CCP1-MET1 intergenic region.
GN YK067M.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Eukaryota; Saccharomycota; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes;
OC NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RA van Vlier-Reedijk J.C., Planta R.J.;
RL Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: STRONG, TO YEAST YB011W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z28292; CA082146.1; -
DR PIR; S38143; S38143.
DR SCD; S0001775; GPT2.
DR GO; GO:0005783; C:cytoplasmic reticulum; IDA.
DR GO; GO:0004366; P:glycerol-3-phosphate O-acetyltransferase acti. .; IDA.
DR GO; GO:0008654; P:phospholipid biosynthesis; IDA.
DR InterPro; IPR002123; Acyltransferase.
DR Pfam; PF01553; Acyltransferase; 1.
DR SMART; SM00563; Plac; 1.
DR Hypothetical protein; Transmembrane.
FT TRANSMEM 31 55 POTENTIAL.
FT TRANSMEM 69 85 POTENTIAL.

```

```

FT TRANSMEM 502 524 POTENTIAL.
FT TRANSMEM 539 555 POTENTIAL.
SQ SEQUENCE 743 AA; 83644 MW; 84B994656B82F15 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 743;
Best Local Similarity 75.0%; Pred. No. 41;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPGXHY 10
Db 294 VVPGXHY 301

RESULT 12
GAAB_METH STANDARD; PRT; 308 AA.
ID GAAB_METH STANDARD; PRT; 308 AA.
AC O26806;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE GMP synthase [glutamine-hydrolyzing] subunit B (BC 6.3.5.2) (GMP
DE synthetase).
GN GUAAB OR MTH710.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OC NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Donette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakey D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Potlter B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics."
RL J. Bacteriol. 179:7135-7155 (1997).
CC -1- CATALYTIC ACTIVITY: ATP + xanthosine 5'-phosphate + L-glutamine +
CC H(2)O = AMP + diphosphate + GMP + L-glutamate.
CC -1- PATHWAY: GMP biosynthesis.
CC -1- SUBUNIT: HETERODIMER COMPOSED OF A GLUTAMINE AMIDOTRANSFERASE
CC SUBUNIT (A) AND A GMP SYNTHASE SUBUNIT (B) (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE GMP SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; A000850; AAB85215.1; ALT_INIT.
DR HSP; P04079; IGPW.
DR HAMAP; MF 00345; -1.
DR InterPro; IPR001674; GMP synt C.
DR Pfam; PF00958; GMP synt C; 1.
DR TIGRfam; TIGR00884; guaA_Cterm; 1.
DR Ligase; GMP biosynthesis; Purine biosynthesis; ATP-binding;
KM DOMAIN 33 184 GMP-BINDING (BY SIMILARITY).
FT NP_BIND 29 35 ATP (BY SIMILARITY).
FT DOMAIN 33 184 GMP-BINDING (BY SIMILARITY).
SQ SEQUENCE 308 AA; 34403 MW; F2DCP6BD20CAEC1 CRC64;

Query Match 61.5%; Score 32; DB 1; Length 308;
Best Local Similarity 63.6%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 EVVPGXHY 11

```

DB 216 SEVESGLHES 226

RESULT 13

ID	ALPB_SHEEP	STANDARD;	PRT;	363 AA.
AC	P52210;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Fructose-bisphosphate aldolase B (EC 4.1.2.13) (Liver-type aldolase).			
GN	ALDOB.			
OS	Ovis aries (Sheep).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Caprinae; Ovis.			
OX	NCBI_TaxID=9940;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Meconephros;			
RX	MEDLINE=9436863; PubMed=8086469;			
RA	Gianquinto L., Palloux E.A., Bezaud J., Serevel N., Kirszenbaum M.,			
RA	Cotinot C.;			
RT	"Cloning and characterization of a full-length cDNA coding for ovine			
RL	aldolase B from fetal meconephros."			
CC	Biochim. Biophys. Acta 1219:223-227(1994).			
CC	-1- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate = glycero-			
CC	phosphate + D-glyceraldehyde 3-phosphate.			
CC	-1- PATHWAY: Glycolysis; sixth step.			
CC	-1- SUBUNIT: Homotetramer (By similarity).			
CC	-1- MISCELLANEOUS: In vertebrates, three forms of this ubiquitous			
CC	glycolytic enzyme are found, aldolase A in muscle, aldolase B in			
CC	liver and aldolase C in brain.			
CC	-1- SIMILARITY: Belongs to class I fructose-bisphosphate aldolase			
CC	family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; 229372; CAA82563.1; -			
DR	PIR; S47540; S47540.			
DR	HSSP; P00883; IADO.			
DR	InterPro; IPR000741; Aldolase_I.			
DR	Pfam; PF00274; Glycolytic_enzy; 1.			
DR	ProDom; PD001128; Aldolase_I; 1.			
DR	PROSITE; PS00158; ALDOLASE_CLASS_I; 1.			
KW	Lyase; Schiff base; Glycolysis; Multigene family.			
FT	INIT MET 0			
FT	BINDING 55			
FT	BINDING 146			
FT	BINDING 229			
FT	ACT_SITE 363			
FT	SEQUENCE 363 AA; 39500 MW; FCB845668212BD CRC64;			
QY	Query Match 61.5%; Score 32; DB 1; Length 363;			
DB	Best Local Similarity 62.5%; Pred. No. 32;			
DB	Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;			
DB	2 EVVPGXK 9			
DB	189 EVIPGSH 196			

RESULT 14

SUOX_DROME

ID SUOX_DROME STANDARD; PRT; 573 AA.

AC	O9WBF4;
DT	28-FEB-2003 (Rel. 41, Created)
DT	28-FEB-2003 (Rel. 41, Last sequence update)
DT	15-SEP-2003 (Rel. 42, Last annotation update)
DE	Probable sulfite oxidase, mitochondrial precursor (EC 1.8.3.1).
GN	CS7280.
OS	Drosophila melanogaster (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
OX	NCBI_TaxID=7227;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Berkeley;
RX	MEDLINE=20196006; PubMed=10731132;
RA	Adams M.D., Celniker S.E., Holt R.A., Byrne C.A., Gocayne J.D.,
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoekling R.A., Galle R.F.,
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA	Burton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA	Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA	Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA	Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA	de Pablos R., Delcher A., Deng Z., Maye A.D., Dew I., Dietz S.M.,
RA	Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA	Dubin K.J., Evangelista C.C., Ferraz C., Ferrelia S., Fleischmann W.,
RA	Forster C., Gabriellian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA	Hoselin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA	Jaisli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA	Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA	Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacלב J.M.,
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA	Reiner K., Remington K., Saunders R.D.C., Scheerer F., Shen H.,
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA	Spieler B., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA	Svaylkes R., Tector C., Turner R., Venter B., Wang A.H., Wang X.,
RA	Wang Z.-Y., Wasmann D.A., Weinstock G.M., Weissenbach J.,
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA	Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA	Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT	"The genome sequence of Drosophila melanogaster."
RL	Science 287:2185-2195(2000).
RP	SEQUENCE FROM N.A.
RC	STRAIN=Berkeley; TISSUE=Embryo;
RC	MEDLINE=22426066; PubMed=12537569;
RX	Stapleton M., Carlson J.M., Brokstein P., Yu C., Champe M.,
RA	George R.A., Guarin H., Krommiller B., Pacלב J.M., Park S., Wan K.H.,
RA	Rubin G.M., Celniker S.E.;
RT	"A Drosophila full-length cDNA resource."
RL	Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC	-1- CATALYTIC ACTIVITY: Sulfite + O(2) + H(2)O = sulfate + H(2)O(2).
CC	-1- COPACATOR: Molybdenum (molybdopterin) and one prothome group (By
CC	similarity).
CC	-1- PATHWAY: Terminal reaction in the oxidative degradation of sulfur-
CC	containing amino acids. It uses cytochrome c as an electron
CC	acceptor.
CC	-1- SUBCELLULAR LOCATION: Mitochondrial intermembrane space (By
CC	similarity).
CC	-1- SIMILARITY: Contains 1 cytochrome b5 heme-binding domain.

CC -1- CAUTION: It is not obvious if the molybdenum-pterin domain is
 CC functional; the conserved cysteine (position 339) is replaced by
 CC Ile-339.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcement/>
 CC or send an email to license@ebi.ac.uk).

DR EMBL; AB003510; AAF8894.1; -
 DR EMBL; AY069352; AAL39497.1; -
 DR HSPB; P07850; 180X.
 DR FLYBase; FBgn0030966; CG7280.
 DR GO; GO:0005758; C:mitochondrial intermembrane space; NAS.
 DR GO; GO:0008482; P:sulfite oxidase activity; NAS.
 DR GO; GO:0030163; P:protein catabolism; NAS.
 DR InterPro; IPR001199; Cyt_B5.
 DR InterPro; IPR000572; Euk_Mb_oxred.
 DR InterPro; IPR005066; Mo-co_dimer.
 DR Pfam; PF00173; heme_1; 1.
 DR Pfam; PF03404; Mo-co_dimer; 1.
 DR Pfam; PF00174; oxidored_molyb; 1.
 DR PRINTS; PR00363; CYTOCHROME B5.
 DR PRINTS; PR00407; EMOPTERIN.
 DR PRODOM; PD000612; Cyt_B5; 1.
 DR PROSITE; PS00191; CYTOCHROME B5_1; 1.
 DR PROSITE; PS0255; CYTOCHROME B5_2; 1.
 DR PROSITE; PS00553; MOLYBDOTERIN_EUK; FALSE NEG.
 DR Hypothetical protein; Oxidoreductase; Mitochondrion; Heme; Molybdenum;
 DR TRANSIT peptide.
 DR TRANSIT 1
 FT CHAIN ? 573 MITOCHONDRION.
 FT DOMAIN 108 186 PROBABLE SULFITE OXIDASE.
 FT DOMAIN 187 206 HEME-BINDING.
 FT DOMAIN 207 573 HINGE (BY SIMILARITY).
 FT METAL 144 144 MOLYBDENUM-PTERIN DOMAIN (BY SIMILARITY).
 FT METAL 144 144 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 FT METAL 168 168 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 FT METAL 287 287 MOLYBDENUM-PTERIN (BY SIMILARITY).
 SO SEQUENCE 573 AA; 64346 MW; 71ABIDB3465D811 CRC64;

Query Match 61.5%; Score 32; DB 1; Length 573;
 Best Local Similarity 50.0%; Pred. No. 51;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Cy 1 BEVVPXGXY 10
 :|:|:|
 503 QEWMPDGRHY 512

DB

RESULT 15
 MUTL_BACSU STANDARD; PRT; 627 AA.
 AC P49850;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA mismatch repair protein mutL.
 GN MUTL.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=96349107; PubMed=8760914;
 RA Gibertini F., Perego M., Albertini A.M., Gallizi A.;
 RT "Bacillus subtilis mutL operon: identification, nucleotide
 sequence and mutagenesis."
 RL Microbiology 142:2021-2029(1996).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogawara N., Moser I., Albertini A.M., Alloni G.,
 RA Auered V., Bertero M.G., Beesleres P., Bolocin A., Borcher S.,
 RA Borris R., Bourrier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruch C.V., Caldwell B., Capiano V., Carter N.M.,
 RA Choi S.K., Codani J.U., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Danisco F., Devine K.M., Duberhoff A., Ehrlich S.D., Emerson P.T.,
 RA Baiton K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizi A., Galleron N.,
 RA Glaser P., Goffeau A., Golightly B.J., Grand G.,
 RA Giuseppe G., Guy B.J., Haga K., Halesch J., Harwood C.R., Henat A.,
 RA Hilbert H., Holsappel S., Hoono S., Hullo M.F., Iaya M., Jones L.,
 RA Joris B., Karamata D., Kasanara Y., Klett-Bianchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Meuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogawa A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Potwollik S., Prescott A.M.,
 RA Peseacan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan B., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccioni D., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Toso V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambuit R., Wedler B., Medler H., Weitzenger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 subtilis.";
 RT Nature 390:249-256(1997).

CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES IN
 CC DNA. IT IS REQUIRED FOR DAM-DEPENDENT METHYL-DIRECTED DNA MISMATCH
 CC REPAIR. MAY ACT AS A "MOLECULAR MATCHMAKER". A PROTEIN THAT
 CC PROMOTES THE FORMATION OF A STABLE COMPLEX BETWEEN TWO OR MORE
 CC DNA-BINDING PROTEINS IN AN ATP-DEPENDENT MANNER WITHOUT ITSELF
 CC BEING PART OF A FINAL EFFECTOR COMPLEX (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTL/HEXB FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcement/>
 CC or send an email to license@ebi.ac.uk).

DR EMBL; U27343; AAB19236.1; -
 DR EMBL; Z99112; CAB13578.1; -
 DR PIR; A69663; A69663.
 DR HSPB; P23367; IBKN.
 DR Subtilist; BG11402; mutL.
 DR HAMAP; MF_00149; -; 1.
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR002099; DNA_mis_repair.
 DR Pfam; PF01119; DNA_mis_repair; 1.
 DR Pfam; PF02518; HATPase_c; 1.
 DR SMART; SM00387; HATPase_c; 1.
 DR TIGRfam; TIGR00585; mutL_1; 1.
 DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
 DR DNA repair; Complete proteome.
 SO SEQUENCE 627 AA; 70431 MW; 068A0509CC265343 CRC64;

Query Match 61.5%; Score 32; DB 1; Length 627;
 Best Local Similarity 54.5%; Pred. No. 56;
 Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Cy 1 BEVVPXGXY 11
 :|:|:|
 488 EMIVPLTFHY 498

Tue Dec 23 09:51:16 2003

us-09-909-164-44.rsp

Page 9

Search completed: December 22, 2003, 17:42:29
Job time : 5.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:27:26 ; Search time 25.2 Seconds

(without alignments)
112.642 Million cell updates/sec

Title: US-09-909-164-44
Perfect score: 52
Sequence: 1 BEVVPXGXHY 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriophage: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	78.8	413	11 Q8K289	Q8K289 mus musculus
2	36	69.2	208	2 Q46486	Q46486 corynebacte
3	36	69.2	1057	16 Q8CPJ4	Q8CPJ4 staphylococ
4	35	67.3	233	10 Q40479	Q40479 nicotiana t
5	35	67.3	237	10 Q9LMS0	Q9LMS0 nicotiana s
6	35	67.3	285	16 Q98H06	Q98H06 thizobium 1
7	35	67.3	308	16 Q9X2E2	Q9X2E2 thermotoga
8	35	67.3	317	9 Q38317	Q38317 lactobacilli
9	35	67.3	360	12 Q27679	Q27679 methanobact
10	35	67.3	678	12 Q9B1X6	Q9B1X6 ceropithec
11	35	67.3	855	5 Q8IME6	Q8IME6 drosophila
12	35	67.3	1028	16 Q8YU11	Q8YU11 bruceella me
13	35	67.3	1044	16 Q8D1H0	Q8D1H0 synecchococ
14	34	65.4	279	16 Q9RXN9	Q9RXN9 deinococcus
15	34	65.4	322	17 Q9H1H8	Q9H1H8 thermoplasm
16	34	65.4	495	5 Q16912	Q16912 caenorhabdi

17	34	65.4	534	17 Q29966	Q29966 archaeoglob
18	34	65.4	822	2 Q9ZHG7	Q9ZHG7 streptococ
19	34	65.4	822	16 Q8RAU1	Q8RAU1 streptococ
20	34	65.4	822	16 Q8D281	Q8D281 streptococ
21	34	65.4	823	16 Q8N282	Q8N282 streptococ
22	34	65.4	823	16 Q8K5Q1	Q8K5Q1 streptococ
23	34	65.4	825	16 Q9XV4	Q9XV4 streptococ
24	34	65.4	962	5 Q8IB11	Q8IB11 plasmodium
25	34	65.4	1442	17 Q96VH5	Q96VH5 sulfolobus
26	33	63.5	46	16 Q8CTU7	Q8CTU7 streptococ
27	33	63.5	78	6 Q9XST4	Q9XST4 canis fami
28	33	63.5	139	2 Q57489	Q57489 bacteroides
29	33	63.5	156	16 Q9PC35	Q9PC35 xylella fas
30	33	63.5	175	6 Q8I033	Q8I033 bos taurus
31	33	63.5	215	6 Q8I031	Q8I031 bos taurus
32	33	63.5	217	4 Q00404	Q00404 homo sapien
33	33	63.5	252	16 Q8YWP1	Q8YWP1 anabaena sp
34	33	63.5	297	6 Q8HX9	Q8HX9 bos taurus
35	33	63.5	299	4 Q9UER9	Q9UER9 homo sapien
36	33	63.5	364	5 Q8IGN5	Q8IGN5 drosophila
37	33	63.5	412	5 Q08523	Q08523 ascaris suu
38	33	63.5	480	5 Q95NKL	Q95NKL plasmodium
39	33	63.5	510	10 Q9SA71	Q9SA71 arabidopsis
40	33	63.5	722	12 Q9QTC7	Q9QTC7 marek's dis
41	33	63.5	748	12 Q9YR01	Q9YR01 ranid herpe
42	33	63.5	802	16 Q9TQM8	Q9TQM8 streptococ
43	33	63.5	816	2 Q9ANT9	Q9ANT9 streptococ
44	33	63.5	819	2 Q9ANT3	Q9ANT3 streptococ
45	33	63.5	819	10 Q9AVK4	Q9AVK4 pisum sativ

ALIGNMENTS

RESULT 1
Q8K289 PRELIMINARY; PRT; 413 AA.
ID Q8K289
AC Q8K289
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Similar to expressed sequence A1987856.
GN 1110004B19R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC032195; AAH32195.1; -
DR MGD; MGI:1915724; 1110004B19R1K.
SQ SEQUENCE 413 AA; 46621 MW; A9BBA1DC70CDA0D5 CRC64;
Query Match 78.8%; Score 41; DB 11; Length 413;
Best Local Similarity 63.6%; Pred. No. 2.1;
Matches 7; Conservativity 1; Mismatches 3; Indels 0; Gaps 0;
Oy 1 BEVVPXGXHY 11
Db 168 BEVVPXGXHY 178
RESULT 2
Q46486 PRELIMINARY; PRT; 208 AA.
ID Q46486
AC Q46486
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE Hypothetical 23.0 kDa protein (Gera).
GN GERA.

OS Corynebacterium xerosis, and
 OS Corynebacterium striatum.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
 OC NCBI_TaxID=1725, 43770;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=C.xerosis; STRAIN=M82B;
 RA MEDLINE=96117603; PubMed=8559800;
 RA Tauch A., Kaasing P., Kallinowski J., Puhler A.;
 RT "The 51,409-bp R-plasmid pTrio from the multiresistant clinical
 RT isolate Corynebacterium striatum M82B is composed of DNA segments
 RT initially identified in soil bacteria and in plant, animal, and human
 RT pathogens";
 RL Mol. Gen. Genet. 263:1-11(2000).
 DR EMBL; U21300; AAC95478.1; -
 DR EMBL; AF024666; AAG03390.1; -
 KM Hypothetical protein; Plasmid.
 SO SEQUENCE 208 AA; 23012 MW; F1504BE1ECDE85A6 CRC64;

Query Match 69.2%; Score 36; DB 2; Length 208;
 Best Local Similarity 50.0%; Pred. No. 12;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVPGXGXY 11
 Db 130 DVPEGRHY 139

RESULT 3
 Q8CPJ4 PRELIMINARY; PRT; 1057 AA.
 AC Q8CPJ4.
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Carbamoyl-phosphate synthase large chain.
 GN SR0879.
 OS Staphylococcus epidermidis.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OC NCBI_TaxID=1282;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 12228;
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
 RA Chen Z., Wen Y.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB016746; AA004476.1; -
 KW Complete proteome.
 SO SEQUENCE 1057 AA; 117391 MW; 8944D7D8DB1CAE59 CRC64;

Query Match 69.2%; Score 36; DB 16; Length 1057;
 Best Local Similarity 63.6%; Pred. No. 71;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVPGXGXY 11
 Db 189 KEVSNGLHY 199

RESULT 4
 Q40479 PRELIMINARY; PRT; 233 AA.
 AC Q40479;

DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE ERBBP-2.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
 OC NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BY4; TISSUE=leaf;
 RA MEDLINE=95276459; PubMed=7756828;
 RA Ome-Takagi M., Shinsai H.;
 RT "Ethylene-inducible DNA binding proteins that interact with an
 RT ethylene responsive element";
 RL Plant Cell 7:173-182(1995).
 DR EMBL; D38125; BAA07324.1; -
 DR HSSP; O80337; 2GCC.
 DR TRANSFAC; T02654; -
 DR InterPro; IPR001471; TF_BRF.
 DR Pfam; PF00847; AP2-domain; 1.
 DR ProDom; PD001423; TF_BRF; 1.
 DR SMART; SM00380; AP2; 1.
 SO SEQUENCE 233 AA; 25563 MW; 6CD16783582C0CB5 CRC64;

Query Match 67.3%; Score 35; DB 10; Length 233;
 Best Local Similarity 60.0%; Pred. No. 23;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVPGXGXY 10
 Db 90 QAVVPGKGRHY 99

RESULT 5
 Q9LW50 PRELIMINARY; PRT; 237 AA.
 AC Q9LW50.
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Ethylene-responsive element binding factor.
 GN NSERF2.
 OS Nicotiana glauca (Wood tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
 OC NCBI_TaxID=4096;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20399450; PubMed=10945353;
 RA Kiteajima S., Koyama T., Ome-Takagi M., Shinsai H., Sato P.;
 RT "Characterization of gene expression of NSERF2, transcription factors
 RT of basic PR genes from Nicotiana glauca".
 RL Plant Cell Physiol. 41:817-824(2000).
 DR EMBL; AB016264; BAA97122.1; -
 DR HSSP; O80337; 2GCC.
 DR InterPro; IPR001471; TF_BRF.
 DR Pfam; PF00847; AP2-domain; 1.
 DR ProDom; PD001423; TF_BRF; 1.
 DR SMART; SM00380; AP2; 1.
 SO SEQUENCE 237 AA; 26243 MW; 01BC3BB51R46298 CRC64;

Query Match 67.3%; Score 35; DB 10; Length 237;
 Best Local Similarity 60.0%; Pred. No. 23;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVPGXGXY 10
 Db 94 QAVVPGKGRHY 103

```

RESULT 6
ID Q98HU6 PRELIMINARY; PRT; 285 AA.
AC Q98HU6;
DT 01-OCT-2001 (TRMBLrel. 18, Created)
DT 01-OCT-2001 (TRMBLrel. 18, Last sequence update)
DT 01-MAR-2003 (TRMBLrel. 23, Last annotation update)
DE Cytochrome c1.
GN ML2705.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFP30309; PubMed=11214968;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Ideawara K., Ishikawa A., Kawashima K., Kimura T.,
RA Klehida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003000; BAB49770.1; -.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR002326; Cyt_C1.
DR Pfam; PF02167; Cytochrome c1; 1.
DR PRINTS; PR00603; CYTOCHROME1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
KW Complete proteome.
SQ SEQUENCE 285 AA; 30961 MW; 3109CDE2711747EE CRC64;

Query Match 67.3%; Score 35; DB 16; Length 285;
Best Local Similarity 55.6%; Pred. No. 2;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

CY 3 VVPGXHY 11
DB 194 VVPGTHY 202

RESULT 7
ID Q9X2E2 PRELIMINARY; PRT; 308 AA.
AC Q9X2E2;
DT 01-NOV-1999 (TRMBLrel. 12, Created)
DT 01-NOV-1999 (TRMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TRMBLrel. 23, Last annotation update)
DE F1SH protease activity modulator HPLK.
GN TM1822.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.B., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback R.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999).
DR EMBL; AB001819; AAD36885.1; -.
DR TIGR; TM1822; -.
DR InterPro; IPR001107; Band 7.
DR InterPro; IPR001972; Stomatol.
DR Pfam; PF01145; Band_7; 1.

```

```

DR PRINTS; PR00721; STOMATIN.
DR SMART; SM00244; PHB; 1.
KW Protease; Complete proteome.
SQ SEQUENCE 308 AA; 34778 MW; AD03603B5101AD CRC64;

Query Match 67.3%; Score 35; DB 16; Length 308;
Best Local Similarity 75.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

CY 3 VVPGXHY 10
DB 41 VVPGTHY 48

RESULT 8
ID Q38317 PRELIMINARY; PRT; 317 AA.
AC Q38317;
DT 01-NOV-1996 (TRMBLrel. 01, Created)
DT 01-NOV-1996 (TRMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TRMBLrel. 23, Last annotation update)
DE Lysin.
GN Lys.
OS Lactobacillus bacteriophage phi adh.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=12417;
RN [1]
RP SEQUENCE FROM N.A.
RA Altermann E.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99384014; PubMed=10452953;
RA Altermann E., Klein J., Heinrich B.;
RT "Primary structure and features of the genome of the Lactobacillus
RT gasseri temperate bacteriophage phi-adh."
RL Gene 236:333-346(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95138034; PubMed=7836107;
RA Heinrich B., Binschofer B., Blesl U.;
RT "Primary structure and functional analysis of the Lys genes of
RT Lactobacillus gasseri bacteriophage phi-adh."
RL J. Bacteriol. 177:723-732(1995).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=99323538; PubMed=8472961;
RA Fremaux C., De Antoni G., Raya R., Klaenhammer T.;
RT "Genetic organization and sequence of the region encoding integrative
RT functions from Lactobacillus gasseri temperate bacteriophage phi-
RT adh."
RL Gene 126:61-66(1993).
RN [5]
RP SEQUENCE FROM N.A.
RA Engel G., Altermann E., Klein J., Heinrich B.;
RT "Structure of a genome region of the Lactobacillus gasseri temperate
RT phage phi adh covering a repressor gene and cognate promoters."
RL Gene 210:67-70(1998).
DR EMBL; AJ11519; CAB52540.1; -.
DR InterPro; IPR002053; Glyco_hydro_25.
DR InterPro; IPR003646; SH3_bac.
DR Pfam; PF01183; Glyco_hydro_25; 1.
DR ProDom; PR004620; Glyco_hydro_25; 1.
DR SMART; SM00641; Glyco_25; 1.
DR SMART; SM00287; SH3; 1.
SQ SEQUENCE 317 AA; 34703 MW; 9FF2715BB43561C7 CRC64;

Query Match 67.3%; Score 35; DB 9; Length 317;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

CY 3 VVPGXHY 11

```

Db 60 VVPMGVHYA 68

RESULT 9

ID 027679 PRELIMINARY; PRT; 360 AA.

AC 027679;

DT 01-JAN-1998 (TReMBLrel. 05, Created)

DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE Cell division protein.

GN MTH1642.

OS Methanobacterium thermoautotrophicum.

OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;

OC Methanobacteriaceae; Methanothermobacter.

NCBI_TaxID=187420;

OK [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Delta H;

RX MEDLINE=98037514; PubMed=9371463;

RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,

RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,

RA Harrison D., Hoang L., Keagle P., Lumm W., Pochler B., Qiu D.,

RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,

RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,

RA Daniels C.J., Mao J.-I., Rice P., Noelling U., Reeve J.N.,

RT "Complete genome sequence of Methanobacterium thermoautotrophicum

RT delah: functional analysis and comparative genomics."

RL J. Bacteriol. 179:7135-7155 (1997).

DR EMBL: AB000923; AAB615.1;

DR InterPro: IPR005140; eRF1_1.

DR InterPro: IPR005141; eRF1_2.

DR InterPro: IPR005142; eRF1_3.

DR InterPro: IPR004405; eRF1.

DR Pfam: PF03463; eRF1_1; 1.

DR Pfam: PF03464; eRF1_2; 1.

DR Pfam: PF03465; eRF1_3; 1.

DR TIGRfam: TIGR00111; pelota; 1.

DR Cell division; Complete proteome.

KW SEQUENCE 360 AA; 40814 MW; 2A000C4B3CEFA69 CRC64;

SQ

Query Match 67.3%; Score 35; DB 17; Length 360;

Best Local Similarity 45.5%; Pred. No. 36;

Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EEVVPXGXY 11

Db 98 EDLVPWGSHT 108

RESULT 10

ID 09B1X6 PRELIMINARY; PRT; 678 AA.

AC 09B1X6;

DT 01-MAR-2001 (TReMBLrel. 16, Created)

DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)

DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)

DE Hypothetical 75.9 kDa protein.

OS Herpesviridae; Herpesvirinae; Herpesviridae.

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Alphaherpesvirinae; Varicellovirinae.

NCBI_TaxID=35245;

OK [1]

RP SEQUENCE FROM N.A.

RA Gray W.L., Starnes H.B., White M.W., Ashburn C.V., Mahalingam R.,

RT "Complete Sequence of the Simian Varicella Virus Genome."

RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF275348; AF27217.1; -.

KW Hypothetical protein.

SQ SEQUENCE 678 AA; 75850 MW; A17B09530512FE3C CRC64;

Query Match 67.3%; Score 35; DB 12; Length 678;

Best Local Similarity 50.0%; Pred. No. 72;

Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EEVVPXGXY 10

Db 147 EEIIPXGTRY 156

RESULT 11

ID 08IME6 PRELIMINARY; PRT; 855 AA.

AC 08IME6;

DT 01-MAR-2003 (TReMBLrel. 23, Created)

DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE CG30437-PC.

GN CG30437.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Sphingidae; Drosophilidae; Drosophila.

NCBI_TaxID=7227;

OK [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yande J.M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Gabor G.L.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Bailew R.M., Baau A., Bakendate J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Berno P.V., Bernier B.P., Bhandal D., Bolshakov S.,

RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,

RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Dou P.L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evaraglis C.C., Ferrara C., Ferreira S., Fleischmann W.,

RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong Y., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.U., Wei M.H., Ibegwan C.,

RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kemsion J.A., Ketchum K.A.,

RA Jallil B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merulov G., Mlechina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,

RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spler B., Spradling A.C., Stapleton M., Strong R., Sun B.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodard J., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,

RA Yeh R.F., Zaveri J.S., Zhan M., Zhang W., Zhou X., Zhu Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,

RT "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195 (2000).

OK [2]

RP SEQUENCE FROM N.A.

RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,

RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA Barton J., An H., Baldwin D., Barton J., Beeson K.Y., Busam D.A.,

RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,

RA Dodson K., Dorsett V., Dou P.L.B., Doyle E., Dresner D., Farfan D.,

RA Ferreira S., Frie B., Galle R.F., Garg N.S., George R.A.,

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,

RA Ibegwan C., Jaitani M., Kruse D., Li P., Mattei B., Moshrefi A.,

McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Munoz J.,
 Pacle J., Parasas V., Park S., Patel S., Pfeiffer B.,
 Phongsavanong S., Plitman G.S., Puri V., Richards S., Scheeler F.,
 Staphleorn M., Strong R., Svirskas R., Tector C., Tyler D.,
 Williams S.M., Zaveri J.S., Smith H.O., Venter C.J., Rubin G.M.;
 "Sequencing of *Drosophila melanogaster* genome."
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [3]
 SEQUENCE FROM N.A.
 RA Miya S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 Hradecky P., Huang Y., Kaminker J.S., Prochnik S.B., Smith C.D.,
 Tupy J.L., Bergman C., Berman B., Carlson J.W., Celisner S.B.,
 Clamp M., Drysdale R., Emmert D., Frise B., de Grey A., Harris N.,
 Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 Searle S.M.J., Smith B., Shu S., Smucnik P., Whitfield B.,
 Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.B.;
 "Annotation of *Drosophila melanogaster* genome."
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [4]
 SEQUENCE FROM N.A.
 RA Adams M.D., Celisner S.B., Gibbs R.A., Rubin G.M., Venter C.J.;
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [5]
 SEQUENCE FROM N.A.
 RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB003786; AA16124.1; - - 39BD5A516D6312DB CRC64;
 SEQUENCE 855 AA; 94532 MW; 39BD5A516D6312DB CRC64;
 Query Match 67.3%; Score 35; DB 5; Length 855;
 Best Local Similarity 66.7%; Pred. No. 93;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 EHVVPXGXH 9
 Db 657 DEVVPSGDH 665
 RESULT 12
 ID Q8YJ11 PRELIMINARY; PRT; 1028 AA.
 AC Q8YJ11;
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE ATP-dependent DNA helicase.
 GN BME10275.
 OS *Brucella melitensis*.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 CC Brucellaceae; Brucella.
 CX NCBI_TaxID=29459;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=16M / ATCC 23456 / Biotype 1;
 RX MEDLINE=20020109; PubMed=11756688;
 RA DelVecchio V.G., Kapetzel V., Redkar R.J., Patra G., Mujar C., Los T.,
 Ivanova N., Anderson I., Bhattacharya A., Lykdis A., Reznik G.,
 Jablonowski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldstein B.,
 Selkov B., Elzer P.H., Hagius S., O'Callaghan D., Lelesson J.-J.,
 Haasekorn R., Kyrides N., Overbeek R.;
 "The genome sequence of the facultative intracellular pathogen
Brucella melitensis."
 RT "The genome sequence of the facultative intracellular pathogen
Brucella melitensis."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 DR EMBL: AB009470; AAL51457.1; - -
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00271; Helicase_C; 1.
 DR SMART: SM00490; HELIC_C; 1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Complete proteome.
 SQ SEQUENCE 1028 AA; 112996 MW; A752B7042572E219 CRC64;

Query Match 67.3%; Score 35; DB 16; Length 1028;
 Best Local Similarity 54.5%; Pred. No. 1,1e+02;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 EHVVPXGXHS 11
 Db 76 EKIVPGRARYS 86
 RESULT 13
 ID Q8DIH0 PRELIMINARY; PRT; 1044 AA.
 AC Q8DIH0;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Multidrug efflux transporter.
 GN TLI1618.
 OS *Synechococcus elongatus* (Thermosynechococcus elongatus).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
 CX NCBI_TaxID=32046;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP-1;
 RX MEDLINE=22225144; PubMed=12240834;
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 Watanabe A., Iriyuchi M., Kawashima K., Kimura T., Kishida Y.,
 Kiyokawa C., Kohara M., Matsumoto M., Takeuchi C., Yamada M., Tabata S.;
 "Complete genome structure of the thermophilic cyanobacterium
Thermosynechococcus elongatus BP-1."
 RT DNA Res. 9:123-130(2002).
 RL EMBL: AP005374; BAC09170.1; - -
 DR EMBL: AP005374; BAC09170.1; - -
 KW Complete proteome.
 SQ SEQUENCE 1044 AA; 113205 MW; 00E9C13F0F63ED2F CRC64;
 Query Match 67.3%; Score 35; DB 16; Length 1044;
 Best Local Similarity 63.6%; Pred. No. 1.2e+02;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 EHVVPXGXHS 11
 Db 843 EHVVPXGXHS 853
 RESULT 14
 ID Q9RXN9 PRELIMINARY; PRT; 279 AA.
 AC Q9RXN9;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Hypothetical protein DR0271.
 GN DR0271.
 OS *Deinococcus radiodurans*.
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
 CC Deinococcaceae; Deinococcus.
 CX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R1;
 RX MEDLINE=2003686; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.P., Hickey E.K., Peterson J.D.,
 Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 McFarr K.S., Qin H., Jiang L., Pamphile M., Crosby M., Shen M.,
 Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
 Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 Fraser C.M.;
 "Genome sequence of the radioresistant bacterium *Deinococcus*
radiodurans R1."
 RT Science 286:1571-1577(1999).
 EMBL: AB001889; AAF09867.1; - -

DR TIGR: DR0271; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 279 AA; 31140 MW; DCEA100E0AEB8831 CRC64;

Query Match 65.4%; Score 34; DB 16; Length 279;
 Best Local Similarity 75.0%; Pred. No. 46;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 VPKXHY 11
 |||
 Db 100 VPLGRHYS 107

RESULT 15

Q9HLH8 PRELIMINARY; PRT; 322 AA.
 AC Q9HLH8;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE Glucose-fructose oxidoreductase related protein.
 GN TA0250.
 OS Thermoplasma acidophilum.
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
 OC Thermoplasmataceae; Thermoplasma.
 OX NCBI_TaxID=2303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 1728;
 RX MEDLINE=20479972; PubMed=11029001;
 RA Ruepp A., Grahl W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
 RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
 RT "The genome sequence of the thermophilic scavenger Thermoplasma
 acidophilum."
 RL Nature 407:508-513(2000).
 DR EMBL: AL445063; CAC1395.1; -;
 DR InterPro: IPR000683; GPO_IDH_MCCA.
 DR Pfam: PF01408; GPO_IDH_MCCA; 1.
 KW Complete proteome.
 SQ SEQUENCE 322 AA; 36918 MW; B8C239E71009D167 CRC64;

Query Match 65.4%; Score 34; DB 17; Length 322;
 Best Local Similarity 75.0%; Pred. No. 53;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 VVPXGXY 10
 |||
 Db 66 VVPDGLHY 73

Search completed: December 22, 2003, 17:51:32
 Job time : 25.2667 secs

PA (CORV-) CORVAS INT INC.
XX L1m-w1lby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
DR
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
PS Claim 17; Page 65; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 94.0%; Score 47; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0096;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 BEVVPXGXQDYS 11
DB 1 BEVVPXGXQDYS 11
RESULT 2
ABB80548
ID ABB80548 standard; peptide; 11 AA.
XX
AC ABB80548;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #28.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
FN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PE 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI L1m-w1lby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C

PT virus protease -
XX
PS Claim 17; Page 65; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 94.0%; Score 47; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0096;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 BEVVPXGXQDYS 11
DB 1 BEVVPXGXQDYS 11
RESULT 3
ABB80551
ID ABB80551 standard; peptide; 11 AA.
XX
AC ABB80551;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #31.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
FN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PE 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI L1m-w1lby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
PS Claim 17; Page 65; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of

CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;

Query Match 94.0%; Score 47; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0096;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGXSDYS 11
|||
DB 1 BEVVPXGXSDYS 11

RESULT 4
ABB80556
ID ABB80556 standard; peptide; 11 AA.
AC ABB80556;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #36.

KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
vtrncide.

XX Synthetic.

OS Location/Qualifiers

FT Key 1
FT Modified-site /note= "N-terminal acetyl"

FT Modified-site 6
FT /note= "N-terminal acetyl"

FT Modified-site 8
FT /note= "N-terminal acetyl"

FT Modified-site 11
FT /note= "D-form residue"

FT Modified-site 11
FT /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

PT Novel peptide compound having hepatitis C virus protease inhibitory

PT activity useful for treating disorders associated with hepatitis C

XX Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC vtrncide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

SQ Sequence 11 AA;

Query Match 94.0%; Score 47; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0096;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGXSDYS 11
|||
DB 1 BEVVPXGXSDYS 11

RESULT 5
ABB80557
ID ABB80557 standard; peptide; 11 AA.
AC ABB80557;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #37.

KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
vtrncide.

XX Synthetic.

OS Location/Qualifiers

FT Key 1
FT Modified-site /note= "N-terminal acetyl"

FT Modified-site 6
FT /note= "N-terminal acetyl"

FT Modified-site 8
FT /note= "N-terminal acetyl"

FT Modified-site 11
FT /note= "D-form residue"

FT Modified-site 11
FT /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

PT Novel peptide compound having hepatitis C virus protease inhibitory

PT activity useful for treating disorders associated with hepatitis C

XX Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC vtrncide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

SQ Sequence 11 AA;

Query Match 94.0%; Score 47; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0096;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGXSDYS 11
|||
DB 1 BEVVPXGXSDYS 11

XX	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #8.
DE	Hepatitis C virus; HCV, serine protease; inhibitor; alpha-ketoamide;
KM	virocide.
XX	Synthetic.
OS	
XX	
FH	Key Location/Qualifiers
FT	Modified-site 1 /note= "N-terminal acetyl"
FT	Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT	Misc-difference 8 /note= "D-form residue"
FT	Modified-site 11 /note= "C-terminal amide"
XX	
PN	MO200208251-A2.
PD	31-JAN-2002.
XX	
PJ	19-JUL-2001; 2001WO-US23169.
PR	21-JUL-2000; 2000US-220101P.
PA	(CORV-) CORVAS INT INC.
PI	Llm-wilby M, Levy OE, Brunck TK,
DR	WPI; 2002-361643/39.
XX	
PT	Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease -
XX	Claim 17, Page 64; 69pp; English.
CC	The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virulence activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus.
SC	Sequence 11 AA;
Query Match	92.0%; Score 46; DB 23; Length 11;
Best Local Similarity	90.9%; Pred. No. 0.015;
Matches 10; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
OY	1 EEVVFXGXDYS 11
Db	1 EEEVPXGMIDYS 11
RESULT 8	
ABB80529	
ID	ABB80529 standard; peptide; 11 AA.
AC	ABB80529;
DT	08-OCT-2002 (first entry)
DS	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.
XX	Hepatitis C virus; HCV, serine protease; inhibitor; alpha-ketoamide;
KM	virocide.
XX	Synthetic.
OS	
XX	

PH Key
FT Modified-site 1 Location/Qualifiers
FT /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
PN W0200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US23169.
XX 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
XX virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 92.0%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015; Mismatches 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
CY 1 BEVVPXGXDYS 11
DB 1 BEVVPXGMDYS 11
RESULT 9
ABB80533
ID ABB80533 standard; peptide; 11 AA.
XX
XX ABB80533;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #13.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT

FT Modified-site 11 /note= "C-terminal amide"
FT
XX W0200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US23169.
XX 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
XX virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 92.0%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015; Mismatches 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
CY 1 BEVVPXGXDYS 11
DB 1 BEVVPXGMDYS 11
RESULT 10
ABB80534
ID ABB80534 standard; peptide; 11 AA.
XX
XX ABB80534;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #14.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
XX W0200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US23169.
XX

XX 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
XX Llm-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus.
XX
XX Sequence 11 AA;
XX

Query Match 92.0%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015; 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EEWVPXGXDYS 11
1 EEWVPXGQDYS 11
DB

RESULT 11
ABB80538
ID ABB80538 standard; peptide; 11 AA.
XX
XX ABB80538;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #18.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 1 /note= "N-terminal acetyl"
XX Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
XX residue 7"
XX
XX Misc-difference 9 /note= "D-form residue"
XX Modified-site 11 /note= "C-terminal amide"
XX
XX W0200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Llm-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX

XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus.
XX
XX Sequence 11 AA;
XX

Query Match 92.0%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015; 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EEWVPXGXDYS 11
1 EEWVPXGQDYS 11
DB

RESULT 12
ABB80542
ID ABB80542 standard; peptide; 11 AA.
XX
XX ABB80542;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #22.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 1 /note= "N-terminal acetyl"
XX Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
XX residue 7"
XX
XX Misc-difference 8 /note= "D-form residue"
XX Modified-site 11 /note= "C-terminal amide"
XX
XX W0200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Llm-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease -
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having

CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 92.0%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
DB 1 EEVVPXGQDYS 11

RESULT 13

AB80543 ID AB80543 standard; peptide; 11 AA.

XX ABB80543;

DT 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #23.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "N-Valyl carbonyl forming keto-amide linkage with
FT residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease -

XX Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 92.0%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
DB 1 EEVVPXGQDYS 11

RESULT 14

AB80561 ID AB80561 standard; peptide; 11 AA.

XX ABB80561;

DT 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #41.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "N-Valyl carbonyl forming keto-amide linkage with
FT residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Modified-site 8 /note= "Oxymethionine"

FT Modified-site 11 /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease -

XX Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

Query Match 92.0%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.015;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGMDYS 11
 |||||
 DB 1 BEVVPXGMDYS 11

Search completed: December 22, 2003, 17:41:02
 Job time : 32.4667 secs

RESULT 15

ABB80562
 ID ABB80562 standard; peptide; 11 AA.

AC ABB80562;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #42.

KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.

OS Synthetic.

Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "N-terminal carbonyl forming keto-amide linkage with
 residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Modified-site 8 /note= "Oxymethionine"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

MO200208251-A2.

31-JAN-2002.

19-JUL-2001; 2001MO-US23169.

21-JUL-2000; 2000US-220101P.

(CORV-) CORVAS INT INC.

Lim-wilby M, Levy OB, Brunck TK;

WPI; 2002-361643/39.

Novel peptide compound having hepatitis C virus protease inhibitory
 activity useful for treating disorders associated with hepatitis C
 virus protease

Claim 17; Page 65; 69pp; English.

The sequence represents a peptide compound of the invention having
 hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 invention are alpha-ketoamide peptide analogues. The peptides have
 virucide activity, and are useful for treating and in the manufacture of
 a medicament to treat disorders associated with HCV protease. A
 pharmaceutical composition comprising the peptide as an active ingredient
 is useful for treating disorders associated with hepatitis C virus.

Sequence 11 AA;

Query Match 92.0%; Score 46; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.015; 1; Indels 0; Gaps 0;
 Matches 10; Conservative 0; Mismatches 1;

QY 1 BEVVPXGMDYS 11
 |||||
 DB 1 BEVVPXGMDYS 11

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 16:37:03 ; Search time 10.133 Seconds
(without alignments)
45.930 Million cell updates/sec

Title: US-09-909-164-45

Sequence: 1 BEVVPXGXDYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	66.0	140	3	US-08-569-147-76
2	33	66.0	140	3	US-08-569-147-82
3	32	64.0	139	4	US-09-134-001C-5124
4	32	64.0	301	4	US-09-830-217-14
5	31	62.0	622	2	US-08-458-146-2
6	31	62.0	622	2	US-08-459-065-2
7	31	62.0	739	3	US-09-413-814-86
8	30.5	61.0	669	4	US-09-617-594A-2
9	30	60.0	59	4	US-08-963-851-14
10	30	60.0	121	4	US-09-153-060-68
11	30	60.0	121	4	US-09-153-060-85
12	30	60.0	122	2	US-08-879-995A-1
13	30	60.0	122	2	US-09-215-096-1
14	30	60.0	135	4	US-09-996-243-359
15	30	60.0	231	3	US-08-926-842B-20
16	30	60.0	240	3	US-08-926-842B-21
17	30	60.0	241	3	US-08-834-776A-2
18	30	60.0	341	3	US-08-853-948B-4
19	30	60.0	348	3	US-08-853-948B-5
20	30	60.0	368	4	US-09-697-367-24
21	30	60.0	421	4	US-09-093-448-4
22	30	60.0	421	4	US-09-813-555-4
23	30	60.0	425	4	US-09-252-991A-25088
24	30	60.0	478	2	US-09-040-799-3
25	30	60.0	478	3	US-09-093-448-1
26	30	60.0	478	3	US-09-093-448-2
27	30	60.0	478	3	US-09-093-448-3

28	30	60.0	478	4	US-09-813-555-1	Sequence 1, Appli
29	30	60.0	478	4	US-09-813-555-2	Sequence 2, Appli
30	30	60.0	478	4	US-09-813-555-3	Sequence 3, Appli
31	30	60.0	480	4	US-09-252-991A-22164	Sequence 22164, A
32	30	60.0	1049	4	US-09-394-272-10	Sequence 10, Appli
33	30	60.0	1068	2	US-08-429-054A-11	Sequence 11, Appli
34	30	60.0	1068	2	US-08-718-777-7	Sequence 7, Appli
35	30	60.0	1068	3	US-09-051-341-7	Sequence 7, Appli
36	30	60.0	1068	3	US-09-394-272-8	Sequence 8, Appli
37	30	60.0	1081	4	US-09-394-272-4	Sequence 4, Appli
38	30	60.0	1083	4	US-09-394-272-11	Sequence 11, Appli
39	30	60.0	1084	4	US-09-394-272-9	Sequence 9, Appli
40	29	58.0	100	2	US-09-047-125-27	Sequence 27, Appli
41	29	58.0	100	3	US-07-736-335B-27	Sequence 27, Appli
42	29	58.0	364	4	US-09-653-375B-2	Sequence 2, Appli
43	29	58.0	381	4	US-09-134-001C-3003	Sequence 3003, Ap
44	29	58.0	493	4	US-09-411-628-10	Sequence 10, Appli
45	29	58.0	543	4	US-09-739-455-4	Sequence 4, Appli

ALIGNMENTS

```

RESULT 1
US-08-569-147-76
Sequence 76, Application US/08569147
Patent No. 6180377
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: HUMANISED ANTIBODIES
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSER: Woodcock Washburn Kurtz Mackiewicz &
ADDRESSER: No. 6180377r1s, LLP
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: U.S.A.
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (ERO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/569,147
FILING DATE: 25-March-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Trujillo, Doreen Yanko
REGISTRATION NUMBER: 35,719
REFERENCE/DOCKET NUMBER: CARP-0047
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 76:
SEQUENCE CHARACTERISTICS:
LENGTH: 140 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-569-147-76
Query Match 66.0%; Score 33; DB 3; Length 140;
Best Local Similarity 75.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

RESULT 2

Q7 3 VVPXGXDY 10
Db 122 VVPXGXDY 129

US-08-569-147-82
; Sequence 82, Application US/08569147
; Patent No. 6180377
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HUMANISED ANTIBODIES
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSER: No. 618037718, LLP
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA: US/08/569,147
; APPLICATION NUMBER: US/08/569,147
; FILING DATE: 25-March-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Trujillo, Doreen Yatho
; REGISTRATION NUMBER: 35,719
; REFERENCE/DOCKET NUMBER: CARP-0047
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3439
; TELEFAX: (215) 568-3100
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-569-147-82

Query Match 66.0%; Score 33; DB 3; Length 140;
Best Local Similarity 75.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 VVPKXGDY 10
DB 122 VVPTGFDY 129
RESULT 3
US-09-134-001C-5124
; Sequence 5124, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5124
; LENGTH: 139
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5124

Query Match 64.0%; Score 32; DB 4; Length 139;
Best Local Similarity 55.6%; Pred. No. 32;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPKXGDYS 11
DB 32 VVPGHDYN 40
RESULT 4
US-09-830-217-14
; Sequence 14, Application US/09830217
; Patent No. 6521441
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Staphylococcus aureus Genes and Polypeptides
; FILE REFERENCE: PB461PCT
; CURRENT APPLICATION NUMBER: US/09/830,217
; CURRENT FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: PCT/US99/06199
; PRIOR FILING DATE: 1999-03-18
; PRIOR APPLICATION NUMBER: 60/080,296
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/084,674
; PRIOR FILING DATE: 1998-05-07
; SOFTWARE: Patentin Ver. 2.0
; NUMBER OF SEQ ID NOS: 22
; SEQ ID NO 14
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-830-217-14

Query Match 64.0%; Score 32; DB 4; Length 301;
Best Local Similarity 55.6%; Pred. No. 77;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 3 VVPKXGDYS 11
DB 195 VVPGRDYN 203
RESULT 5
US-08-459-146-2
; Sequence 2, Application US/08459146
; Patent No. 5866405
; GENERAL INFORMATION:
; APPLICANT: Choi, Gil Ho
; APPLICANT: Nuss, Donald Lee
; TITLE OF INVENTION: Genetically Engineered Transmissible
; TITLE OF INVENTION: Hypovirulence
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: George W. Gould, Esq., Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/459,146
; APPLICATION NUMBER: US/08/459,146
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 07/832,117
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Roseman, Catherine R
; REGISTRATION NUMBER: 34,240
; REFERENCE/DOCKET NUMBER: 8589
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (201) 235-6208
TELEFAX: (201) 235-3500
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 622 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULAR TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Endocytia parasitica (Cryphonectria
ORGANISM: parasitica)
STRAIN: EP713
US-08-459-146-2

Query Match 62.0%; Score 31; DB 2; Length 622;
Best Local Similarity 85.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 BEVVPXG 7
Db 31 BEVVPAG 37

RESULT 6
US-08-459-065-2
Sequence 2, Application US/08459065
Patent No. 5882642
GENERAL INFORMATION:
APPLICANT: Choi, Gil Ho
APPLICANT: Nuss, Donald Lee
TITLE OF INVENTION: Genetically Engineered Transmissible
TITLE OF INVENTION: Hypovirulence
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.
STREET: 340 Kingland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,065
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/832,117
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Roseman, Catherine R
REGISTRATION NUMBER: 34,240
REFERENCE/DOCKET NUMBER: 8589
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-6208
TELEFAX: (201) 235-3500
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 622 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULAR TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Endocytia parasitica (Cryphonectria
ORGANISM: parasitica)
STRAIN: EP713
US-08-459-065-2

Query Match 62.0%; Score 31; DB 2; Length 622;
Best Local Similarity 85.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 BEVVPXG 7
Db 31 BEVVPAG 37

RESULT 7
US-09-413-814-86
Sequence 86, Application US/09413814
Patent No. 6225064
GENERAL INFORMATION:
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
APPLICANT: Birexol-Myers Squibb, Co.
APPLICANT: Beyer, Stefan
APPLICANT: Bloeker, Helmut
APPLICANT: Brandt, Petra
APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hoile, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
TITLE OF INVENTION: heteropolyketide compounds
FILE REFERENCE: PCT/US 99/2535
CURRENT APPLICATION NUMBER: US/09/413,814
CURRENT FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
EARLIER FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 107
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 86
LENGTH: 739
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-86

Query Match 62.0%; Score 31; DB 3; Length 739;
Best Local Similarity 62.5%; Pred. No. 3.4e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 VPXGXDYS 11
Db 663 IPLGXDYS 670

RESULT 8
US-09-617-594A-2
Sequence 2, Application US/09617594A
Patent No. 6541458
GENERAL INFORMATION:
APPLICANT: Audonnet, et al.
TITLE OF INVENTION: FELINE CALICIVIRUS GENES AND VACCINES IN PARTICULAR RECOMBINANT
FILE REFERENCE: 454313-3151.1
CURRENT APPLICATION NUMBER: US/09/617,594A
CURRENT FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/193,332
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: France 00 01761
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: France 99 09421
PRIOR FILING DATE: 1999-07-16
NUMBER OF SEQ ID NOS: 26
SOFTWARE: Patentin Version 3.0
SEQ ID NO 2
LENGTH: 669
TYPE: PRT
ORGANISM: Feline calicivirus
US-09-617-594A-2

Query Match 61.0%; Score 30.5; DB 4; Length 669;
Best Local Similarity 63.6%; Pred. No. 3.8e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 BEVVPXGDYS 11
DB 428 BEVTPAG-DYS 437

RESULT 9

US-08-963-851-14
Sequence 14, Application US/08963851
Patent No. 6300116
GENERAL INFORMATION:
APPLICANT: VAN DER OSTEN, CLAUS
APPLICANT: HALKIER, TORDEN
APPLICANT: ANDERSEN, CARSTEN
APPLICANT: BAUDITZ, PETER
APPLICANT: HANSEN, PETER KAMP
TITLE OF INVENTION: PROTEASE VARIANTS AND COMPOSITIONS
FILE REFERENCE: 4946,200-US
CURRENT APPLICATION NUMBER: US/08/963,851
CURRENT FILING DATE: 1997-11-04
NUMBER OF SEQ ID NOS: 35
SOFTWARE: FASTSQ For Windows Version 3.0
SEQ ID NO 14
LENGTH: 59
TYPE: PRT
ORGANISM: Enterococcus faecalis
US-08-963-851-14

Query Match 60.0%; Score 30; DB 4; Length 59;
Best Local Similarity 45.5%; Pred. No. 31;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGDYS 11
DB 38 EKHIPGLGYS 48

RESULT 10

US-09-152-060-68
Sequence 68, Application US/09152060
Patent No. 6448230
GENERAL INFORMATION:
APPLICANT: ROSEN ET AL.
TITLE OF INVENTION: 28 Human Secreted Proteins
FILE REFERENCE: P2003PI.US
CURRENT APPLICATION NUMBER: US/09/152,060
CURRENT FILING DATE: 1998-09-11
EARLIER APPLICATION NUMBER: PCT/US98/04858
EARLIER FILING DATE: 1998-03-12
EARLIER APPLICATION NUMBER: 60/040,762
EARLIER FILING DATE: 1997-03-14
EARLIER APPLICATION NUMBER: 60/040,710
EARLIER FILING DATE: 1997-03-14
EARLIER APPLICATION NUMBER: 60/050,934
EARLIER FILING DATE: 1997-05-30
EARLIER APPLICATION NUMBER: 60/048,189
EARLIER FILING DATE: 1997-05-30
EARLIER APPLICATION NUMBER: 60/048,100
EARLIER FILING DATE: 1997-05-30
EARLIER APPLICATION NUMBER: 60/048,357
EARLIER FILING DATE: 1997-05-30
EARLIER APPLICATION NUMBER: 60/048,189
EARLIER FILING DATE: 1997-05-30
EARLIER APPLICATION NUMBER: 60/057,765
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/048,970
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/068,368
EARLIER FILING DATE: 1997-12-19
NUMBER OF SEQ ID NOS: 118
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 68

LENGTH: 121
TYPE: PRT
ORGANISM: Homo sapiens
US-09-152-060-68

Query Match 60.0%; Score 30; DB 4; Length 121;
Best Local Similarity 85.7%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXG 7
DB 28 BEVVPBG 34

RESULT 11

US-09-152-060-85
Sequence 85, Application US/09152060
Patent No. 6448230
GENERAL INFORMATION:
APPLICANT: ROSEN ET AL.
TITLE OF INVENTION: 28 Human Secreted Proteins
FILE REFERENCE: P2003PI.US
CURRENT APPLICATION NUMBER: US/09/152,060
CURRENT FILING DATE: 1998-09-11
EARLIER APPLICATION NUMBER: PCT/US98/04858
EARLIER FILING DATE: 1998-03-12
EARLIER APPLICATION NUMBER: 60/040,762
EARLIER FILING DATE: 1997-03-14
EARLIER APPLICATION NUMBER: 60/040,710
EARLIER FILING DATE: 1997-03-14
EARLIER APPLICATION NUMBER: 60/050,934
EARLIER FILING DATE: 1997-05-30
EARLIER APPLICATION NUMBER: 60/048,100
EARLIER FILING DATE: 1997-05-30
EARLIER APPLICATION NUMBER: 60/048,357
EARLIER FILING DATE: 1997-05-30
EARLIER APPLICATION NUMBER: 60/048,189
EARLIER FILING DATE: 1997-05-30
EARLIER APPLICATION NUMBER: 60/057,765
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/048,970
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/068,368
EARLIER FILING DATE: 1997-12-19
NUMBER OF SEQ ID NOS: 118
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 85
LENGTH: 121
TYPE: PRT
ORGANISM: Homo sapiens
FEATURES:
NAME/KEY: SITE
LOCATION: (67)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (89)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-152-060-85

Query Match 60.0%; Score 30; DB 4; Length 121;
Best Local Similarity 85.7%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXG 7
DB 28 BEVVPBG 34

RESULT 12

US-08-879-995A-1
Sequence 1, Application US/08879995A
Patent No. 5985606

GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Kaseer, Matthew R.
TITLE OF INVENTION: HUMAN PREPROTACHYKININ B
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/879,995A
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0326 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 122 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRAITTU03
CLONE: 2109906
US-08-879-995A-1

Query Match 60.0%; Score 30; DB 2; Length 122;
Best Local Similarity 85.7%; Pred. No. 70;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXG 7
DB 28 BEVVPXG 34

RESULT 13
US-09-215-096-1
Sequence 1, Application US/09215096
Patent No. 6008194
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Kaseer, Matthew R.
TITLE OF INVENTION: HUMAN PREPROTACHYKININ B
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/215,096
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/879,995
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0326 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 122 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRAITTU03
CLONE: 2109906
US-09-215-096-1

Query Match 60.0%; Score 30; DB 3; Length 122;
Best Local Similarity 85.7%; Pred. No. 70;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXG 7
DB 28 BEVVPXG 34

RESULT 14
US-09-996-243-359
Sequence 359, Application US/09996243
Patent No. 6478825
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Batton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Pong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Guiney, Austin L.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas P.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
Acids Encoding the Same
FILER REFERENCE: P2730P1C13
CURRENT APPLICATION NUMBER: US/09/996,243
CURRENT FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16

; PRIOR APPLICATION NUMBER: 60/090863
 ; PRIOR FILING DATE: 1998-06-26
 ; PRIOR APPLICATION NUMBER: 60/091360
 ; PRIOR FILING DATE: 1998-07-01
 ; PRIOR APPLICATION NUMBER: 60/091478
 ; PRIOR FILING DATE: 1998-07-02
 ; PRIOR APPLICATION NUMBER: 60/091544
 ; PRIOR FILING DATE: 1998-07-01
 ; PRIOR APPLICATION NUMBER: 60/091519
 ; PRIOR FILING DATE: 1998-07-02
 ; PRIOR APPLICATION NUMBER: 60/091626
 ; PRIOR FILING DATE: 1998-07-02
 ; PRIOR APPLICATION NUMBER: 60/091633
 ; PRIOR FILING DATE: 1998-07-02
 ; PRIOR APPLICATION NUMBER: 60/091978
 ; PRIOR FILING DATE: 1998-07-07
 ; PRIOR APPLICATION NUMBER: 60/091982
 ; PRIOR FILING DATE: 1998-07-07
 ; PRIOR APPLICATION NUMBER: 60/092182
 ; PRIOR FILING DATE: 1998-07-09

Query Match 60.0%; Score 30; DB 4; Length 135;
 Best Local Similarity 85.7%; Pred. No. 78;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVYPXG 7
 |||||
 Db 28 EEVYPXG 34

RESULT 15
 US-08-926-842B-20
 ; Sequence 20, Application US/08926842B
 ; Patent No. 6030807
 ; GENERAL INFORMATION:
 ; APPLICANT: Sa-No. 6030807neira, Isabel
 ; TITLE OF INVENTION: HIGHLY REGULABLE PROMOTER FOR HETEROLOGOUS GENE
 ; NUMBER OF SEQUENCES: 64
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSER: Klauber & Jackson
 ; STREET: 411 Hackensack Avenue
 ; CITY: Hackensack
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07601
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/926,842B
 ; FILING DATE: 10-SEP-1997
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jackson Bag., David A.
 ; REGISTRATION NUMBER: 26,742
 ; REFERENCE/DOCKET NUMBER: 600-1-089 N
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201 487-5800
 ; TELEFAX: 201 343-1684
 ; TELEX: 133521
 ; INFORMATION FOR SEQ ID NO: 20:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 231 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULAR TYPE: Protein
 ; HYPOTHEICAL: NO
 ; ORIGINAL SOURCE:

; ORGANISM: Escherichia coli
 US-08-926-842B-20

Query Match 60.0%; Score 30; DB 3; Length 231;
 Best Local Similarity 55.6%; Pred. No. 1,4e+02;
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 VPRGXDXS 11
 : |||||
 Db 41 IKPSGVDS 49

Search completed: December 22, 2003, 16:43:46
 Job time : 11.1333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:40:14 ; Search time 20.8667 Seconds

(without alignments)
98.451 Million cell updates/sec

Title: US-09-909-164-45
Perfect score: 50
Sequence: 1 EVVXPXGXDYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgnt2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgnt2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgnt2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgnt2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgnt2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgnt2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgnt2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgnt2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgnt2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgnt2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgnt2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgnt2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgnt2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgnt2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgnt2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgnt2_6/ptodata/2/pubpaa/US10C_NEW_PUB.pep.*
17: /cgnt2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgnt2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	34	68.0	283	10 US-09-738-626-4881	Sequence 4881, App1
2	34	68.0	322	15 US-10-156-761-12418	Sequence 12418, A
3	33	66.0	153	12 US-10-029-386-32486	Sequence 32486, A
4	33	66.0	222	12 US-10-091-007-58	Sequence 58, App1
5	33	66.0	440	11 US-09-813-408-27	Sequence 27, App1
6	32	64.0	14	12 US-10-231-417-403	Sequence 403, App
7	32	64.0	14	12 US-10-231-417-406	Sequence 406, App
8	32	64.0	301	12 US-10-278-946-14	Sequence 14, App1
9	32	64.0	873	12 US-10-132-350-2	Sequence 2, App1
10	32	64.0	873	12 US-10-132-350-4	Sequence 4, App1
11	32	64.0	1163	12 US-10-122-067-4	Sequence 4, App1
12	32	64.0	1499	12 US-10-096-534-67	Sequence 67, App1
13	32	64.0	1499	12 US-10-122-067-2	Sequence 2, App1
14	32	64.0	3472	12 US-10-029-120-4	Sequence 4, App1
15	32	64.0	3472	14 US-10-027-806-4	Sequence 4, App1

16	32	64.0	3472	14 US-10-034-623-4	Sequence 4, App1
17	32	64.0	3472	15 US-10-027-801-4	Sequence 4, App1
18	31	62.0	299	9 US-09-815-242-10597	Sequence 10597, A
19	31	62.0	653	11 US-09-820-843A-26	Sequence 26, App1
20	31	62.0	846	9 US-09-815-242-13904	Sequence 13904, A
21	30.5	61.0	669	15 US-10-209-507-2	Sequence 2, App1
22	30	60.0	7	11 US-09-909-062-1	Sequence 1, App1
23	30	60.0	7	11 US-09-909-062-1	Sequence 9, App1
24	30	60.0	7	11 US-09-909-062-130	Sequence 130, App
25	30	60.0	11	10 US-09-943-123-22	Sequence 22, App1
26	30	60.0	59	10 US-09-948-080-14	Sequence 14, App1
27	30	60.0	79	11 US-09-764-891-4601	Sequence 4601, App
28	30	60.0	121	9 US-09-853-161-85	Sequence 68, App1
29	30	60.0	121	9 US-09-853-161-85	Sequence 85, App1
30	30	60.0	121	9 US-09-852-659A-68	Sequence 68, App1
31	30	60.0	121	9 US-09-852-659A-85	Sequence 85, App1
32	30	60.0	121	10 US-09-852-797-85	Sequence 85, App1
33	30	60.0	121	10 US-09-852-797-85	Sequence 85, App1
34	30	60.0	126	12 US-10-193-840-946	Sequence 946, App
35	30	60.0	135	9 US-09-989-723-359	Sequence 359, App
36	30	60.0	135	9 US-09-989-723-359	Sequence 359, App
37	30	60.0	135	9 US-09-989-723-359	Sequence 359, App
38	30	60.0	135	9 US-09-989-723-359	Sequence 359, App
39	30	60.0	135	10 US-09-989-723-359	Sequence 359, App
40	30	60.0	135	10 US-09-989-723-359	Sequence 359, App
41	30	60.0	135	10 US-09-991-073-359	Sequence 359, App
42	30	60.0	135	10 US-09-990-442-359	Sequence 359, App
43	30	60.0	135	10 US-09-991-163-359	Sequence 359, App
44	30	60.0	135	10 US-09-993-604-359	Sequence 359, App
45	30	60.0	135	10 US-09-990-456-359	Sequence 359, App

ALIGNMENTS

RESULT 1
US-09-738-626-4881
Sequence 4881, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738, 626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 4881
LENGTH: 283
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-4881

Query Match 68.0%; Score 34; DB 10; Length 283;
Best Local Similarity 75.0%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 4 VPXGXDYS 11

Db 56 VEPAGADYS 63

RESULT 2

US-10-156-761-12418
Sequence 12418, Application US/10156761
Publication No. US20030118018A1
GENERAL INFORMATION:
APPLICANT: ODMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBATA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 12418
LENGTH: 322
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-12418

Query Match 68.0%; Score 34; DB 15; Length 322;

Best Local Similarity 66.7%; Pred. No. 61;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EEVVPXGXD 9

Db 65 EDVVPAGAD 73

RESULT 3

US-10-029-386-32486
Sequence 32486, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
FILE REFERENCE: A60MICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 32486
LENGTH: 153
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AF002994.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
OTHER INFORMATION: SWISSPROT HIT: P43363, EVALUATE 5.00e-46
US-10-029-386-32486

Query Match 66.0%; Score 33; DB 12; Length 153;
Best Local Similarity 60.0%; Pred. No. 43;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EEVVPXGXDY 10
Db 15 KEVVPFGHSY 24

RESULT 4

US-10-091-007-58
Sequence 58, Application US/10091007
Publication No. US20030170782A1
GENERAL INFORMATION:
APPLICANT: Microbial Techniques limited
APPLICANT: Le Page, Richard W F Wells, Jeremy M
APPLICANT: Hamblify, Sean B
TITLE OF INVENTION: Proteins
FILE REFERENCE: PRC/P21978MO
CURRENT APPLICATION NUMBER: US/10/091,007
CURRENT FILING DATE: 2002-03-06
PRIOR APPLICATION NUMBER: GB 9921125.2
PRIOR FILING DATE: 1999-09-07
NUMBER OF SEQ ID NOS: 276
SOFTWARE: Patent version 3.0
SEQ ID NO 58
LENGTH: 222
TYPE: PRT
ORGANISM: Streptococcus agalactiae
US-10-091-007-58

Query Match 66.0%; Score 33; DB 12; Length 222;

Best Local Similarity 50.0%; Pred. No. 65;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EEVVPXGXDY 10

Db 201 KXVVPFGDY 210

RESULT 5

US-09-813-408-27
Sequence 27, Application US/09813408
Publication No. US2003049619A1
GENERAL INFORMATION:
APPLICANT: Delagrave, Simon
APPLICANT: Marrs, Barry
TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides And Combinatorial
FILE REFERENCE: HER0041
CURRENT APPLICATION NUMBER: US/09/813,408
CURRENT FILING DATE: 2001-03-21
NUMBER OF SEQ ID NOS: 85
SOFTWARE: Patent version 3.0
SEQ ID NO 27
LENGTH: 440
TYPE: PRT
ORGANISM: Aeropyrum pernix
US-09-813-408-27

Query Match 66.0%; Score 33; DB 11; Length 440;

Best Local Similarity 66.7%; Pred. No. 1,4e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EEVVPXGXDY 10

Db 120 EALPFGVDY 128

RESULT 6

US-10-231-417-403
Sequence 403, Application US/100231417
Publication No. US20030176681A1
GENERAL INFORMATION:
APPLICANT: Peng et al.
TITLE OF INVENTION: 148 Human Secreted Proteins
FILE REFERENCE: P2019P1

Query Match 66.0%; Score 33; DB 11; Length 440;

Best Local Similarity 66.7%; Pred. No. 1,4e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EEVVPXGXDY 10

Db 120 EALPFGVDY 128

RESULT 6

US-10-231-417-403
Sequence 403, Application US/100231417
Publication No. US20030176681A1
GENERAL INFORMATION:
APPLICANT: Peng et al.
TITLE OF INVENTION: 148 Human Secreted Proteins
FILE REFERENCE: P2019P1

```
; CURRENT APPLICATION NUMBER: US/10/231,417
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: US/09/296,622
; PRIOR FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 619
; SOFTWARE: Patentn Ver. 2.0
; SEQ ID NO 403
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-231-417-403

Query Match      64.0%; Score 32; DB 12; Length 14;
Best Local Similarity 75.0%; Pred. No. 4.9;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      4 VPKGXDYS 11
      ||| |||
Db      1 VPCGTDYS 8

RESULT 7
US-10-231-417-406
; Sequence 406, Application US/10231417
; Publication No. US20030176681A1
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; TITLE OF INVENTION: 148 Human Secreted Proteins
; FILE REFERENCE: P2019P1
; CURRENT APPLICATION NUMBER: US/10/231,417
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: US/09/296,622
; PRIOR FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 619
; SOFTWARE: Patentn Ver. 2.0
; SEQ ID NO 406
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-231-417-406

Query Match      64.0%; Score 32; DB 12; Length 14;
Best Local Similarity 75.0%; Pred. No. 4.9;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      4 VPKGXDYS 11
      ||| |||
Db      1 VPCGTDYS 8

RESULT 8
US-10-278-946-14
; Sequence 14, Application US/10278946
; Publication No. US20030153733A1
; GENERAL INFORMATION:
; APPLICANT: Simpson et al.
; TITLE OF INVENTION: Staphylococcus aureus Genes and Polypeptides
; FILE REFERENCE: PB61USDI
; CURRENT APPLICATION NUMBER: US/10/278,946
; CURRENT FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: US 09/830,217
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: PCT/US99/06199
; PRIOR FILING DATE: 1999-03-18
; PRIOR APPLICATION NUMBER: US 60/078,682
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: US 60/080,296
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: US 60/084,674
; PRIOR FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentn version 3.1
; SEQ ID NO 14
```

```
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-278-946-14

Query Match      64.0%; Score 32; DB 12; Length 301;
Best Local Similarity 55.6%; Pred. No. 1.5e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy      3 VVPKXGDYS 11
      ||| |||
Db      195 IVPKGRDYN 203

RESULT 9
US-10-132-350-2
; Sequence 2, Application US/10132350
; Publication No. US20030166855A1
; GENERAL INFORMATION:
; APPLICANT: Acevedo, Pedro A. Navarro
; APPLICANT: Duwick, Jonathan P.
; APPLICANT: Kolomiets, Mikhailo V.
; APPLICANT: Simmons, Carl R.
; TITLE OF INVENTION: Lipoxigenase Polynucleotides and Methods
; FILE REFERENCE: 35718/246439
; CURRENT APPLICATION NUMBER: US/10/132,350
; CURRENT FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: US 60/286,889
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/305,366
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 873
; TYPE: PRT
; ORGANISM: Zea mays
US-10-132-350-2

Query Match      64.0%; Score 32; DB 12; Length 873;
Best Local Similarity 50.0%; Pred. No. 4.8e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy      1 BEVVPKXGDY 10
      ::||| |||
Db      340 KDVLFPVGADY 349

RESULT 10
US-10-132-350-4
; Sequence 4, Application US/10132350
; Publication No. US20030166855A1
; GENERAL INFORMATION:
; APPLICANT: Acevedo, Pedro A. Navarro
; APPLICANT: Duwick, Jonathan P.
; APPLICANT: Kolomiets, Mikhailo V.
; APPLICANT: Simmons, Carl R.
; TITLE OF INVENTION: Lipoxigenase Polynucleotides and Methods
; FILE REFERENCE: 35718/246439
; CURRENT APPLICATION NUMBER: US/10/132,350
; CURRENT FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: US 60/286,889
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/305,366
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 873
; TYPE: PRT
; ORGANISM: Zea mays
```

US-10-132-350-4

Query Match 64.0%; Score 32; DB 12; Length 873;
Best Local Similarity 50.0%; Pred. No. 4.8e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 10
: : : : :
DB 340 KDVLPGVGDY 349

RESULT 11

US-10-122-067-4
; Sequence 4, Application US/10122067
; Publication No. US20030165883A1
; GENERAL INFORMATION:
; APPLICANT: RORY A.J. CURTIS
; TITLE OF INVENTION: 27091, A PHOSPHOLIPID TRANSPORTING
; FILE REFERENCE: MP101-058PIR
; CURRENT APPLICATION NUMBER: US/10/122,067
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/283434
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1163
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-122-067-4

Query Match 64.0%; Score 32; DB 12; Length 1163;
Best Local Similarity 63.6%; Pred. No. 6.5e+02;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
: : : : :
DB 133 EEVVPRGGSVS 143

RESULT 12
US-10-096-534-67
; Sequence 67, Application US/10096534
; Publication No. US20030166887A1
; GENERAL INFORMATION:
; APPLICANT: The Brigham and Women's Hospital, Inc.
; APPLICANT: Yates, Karen
; APPLICANT: Mizuno, Shuichi
; APPLICANT: Glowacki, Julie
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF SKELETAL DEGENERATION CONDITIONS
; FILE REFERENCE: B0801/7244/KX/ERP
; CURRENT APPLICATION NUMBER: US/10/096,534
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: US 60/274,980
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 67
; LENGTH: 1499
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-096-534-67

Query Match 64.0%; Score 32; DB 12; Length 1499;
Best Local Similarity 63.6%; Pred. No. 8.7e+02;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
: : : : :
DB 469 EEVVPRGGSVS 479

RESULT 13

US-10-122-067-2
; Sequence 2, Application US/10122067
; Publication No. US20030165883A1
; GENERAL INFORMATION:
; APPLICANT: RORY A.J. CURTIS
; TITLE OF INVENTION: 27091, A PHOSPHOLIPID TRANSPORTING
; FILE REFERENCE: MP101-058PIR
; CURRENT APPLICATION NUMBER: US/10/122,067
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/283434
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1499
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-122-067-2

Query Match 64.0%; Score 32; DB 12; Length 1499;
Best Local Similarity 63.6%; Pred. No. 8.7e+02;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
: : : : :
DB 469 EEVVPRGGSVS 479

RESULT 14

US-10-029-120-4
; Sequence 4, Application US/10029120
; Publication No. US20030175708A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schlieper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM
; FILE REFERENCE: DCOF.002A
; CURRENT APPLICATION NUMBER: US/10/029,120
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-029-120-4

Query Match 64.0%; Score 32; DB 12; Length 3472;
Best Local Similarity 45.5%; Pred. No. 2.2e+03;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
: : : : :
DB 2294 EDVLPGRISFS 2304

RESULT 15

US-10-027-806-4
; Sequence 4, Application US/10027806
; Publication No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schlieper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM
; FILE REFERENCE: DCOF.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; PRIOR FILING DATE: 2001-12-21

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
 ; NUMBER OF SEQ ID NOS: 123
 ; SOFTWARE: FASTSEQ for Windows Version 3.0
 ; SEQ ID NO 4
 ; LENGTH: 3472
 ; TYPE: PRT
 ; ORGANISM: Cenarchaeum symbiosum
 US-10-027-806-4

Query Match 64.0%; Score 32; DB 14; Length 3472;
 Best Local Similarity 45.5%; Pred. No. 2.2e+03;
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Cy 1 BRVVPXGXDYS 11
 Db 2294 EDV1PRGISFS 2304

Search completed: December 22, 2003, 17:32:45
 Job time : 20.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 17:24:36 ; Search time 9.06667 Seconds

(without alignments)
116.675 Million cell updates/sec

Title: US-09-909-164-45

Perfect score: 50

Sequence: 1 BEVVPXGXDXS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	36	72.0	363	2	D69551
2	35	70.0	290	2	AG3104
3	35	70.0	290	2	D98182
4	35	70.0	587	2	P81138
5	35	70.0	906	2	T48898
6	34	70.0	908	2	T48899
7	34	68.0	102	2	A42452
8	34	68.0	156	2	S54619
9	34	68.0	247	2	A96001
10	34	68.0	257	2	A96546
11	34	68.0	354	2	P82491
12	34	68.0	433	2	H87660
13	34	68.0	2747	2	B49132
14	33	66.0	124	1	VUJ81
15	33	66.0	165	2	AG1272
16	33	66.0	165	2	AH1635
17	33	66.0	196	2	AD0454
18	33	66.0	225	2	S57810
19	33	66.0	327	2	S40753
20	33	66.0	421	1	DERTCM
21	33	66.0	440	2	H72784
22	33	66.0	1028	2	AP3286
23	33	66.0	1088	2	D82246
24	33	66.0	1150	2	T20173
25	32	64.0	99	2	S60210
26	32	64.0	155	2	S38255
27	32	64.0	168	2	S58208
28	32	64.0	196	2	AI0931
29	32	64.0	301	2	P89957

ALIGNMENTS

30	32	64.0	307	2	P84330	hypothetical prote
31	32	64.0	314	2	AH1912	hypothetical prote
32	32	64.0	357	1	G69290	probable hexosyltr
33	32	64.0	366	2	G69350	L-lactate dehydrog
34	32	64.0	425	2	T24111	hypothetical prote
35	32	64.0	427	2	F64064	colB protein - Hae
36	32	64.0	565	2	B86665	ABC transporter AT
37	32	64.0	632	2	H84350	oligopeptidase (im
38	32	64.0	672	2	G88651	protein B0212.3 (1
39	32	64.0	1474	2	F69009	probable membrane
40	32	64.0	3472	2	T31308	hypothetical 367K
41	32	64.0	6658	2	T13931	proteoglycan - fruit
42	31	62.0	117	2	A69487	response regulator
43	31	62.0	202	2	H97247	S-adenosylmethioni
44	31	62.0	233	2	E72330	conserved hypochet
45	31	62.0	296	2	F72745	hypothetical prote

RESULT 1
D69551
Conserved hypothetical protein AP2411 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: D69551
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dode
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dode
G:Glock, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
N:ature 390, 364-370, 1997
A:Authors: Overbeek, R.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Moese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archa
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: D69551
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-363 <KLB>
A:Cross-references: GB:AE001109; GB:AE00782; NID:G2689432; PIDN:AA91255.1; PID:G2650

Query Match 72.0%; Score 36; DB 2; Length 363;
Best Local Similarity 54.5%; Pred. No. 9.1;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Cy 1 BEVVPXGXDXS 11
Db 120 ENIVPGIDRS 130

RESULT 2
AG3104
6-O-methylguanine-DNA methyltransferase [imported] - Agrobacterium tumefaciens (strain
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AG3104
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo,
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; McCle
science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm
ster, R.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AG3104
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-290 <KUR>
A:Cross-references: GB:AE008689; PIDN:AA145253.1; PID:G17742937; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: ada
A:Map position: linear chromosome

Query Match 70.0%; Score 35; DB 2; Length 290;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
OY 1 EBYVPXGXDY 10
DB 9 EDITPISGDY 18

RESULT 3
D98182
O6-methylguanine-DNA methyltransferase PA2118 [imported] - Agrobacterium tumefaciens (st
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: D98182
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Gurolo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: D98182
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-290 <KOR>
A:Cross-references: GB:AE007870; PIDN:AAK8982.1; PID:G15158766; GSPDB:GN00170
A:Gene: AGR L 818
A:Map position: linear chromosome

Query Match 70.0%; Score 35; DB 2; Length 290;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
OY 1 EBYVPXGXDY 10
DB 9 EDITPISGDY 18

RESULT 4
F81138
succinate dehydrogenase, flavoprotein chain NMB0950 [imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: F81138
R:Teitelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignan, V.; Pizsa, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: F81138
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-587 <TET>
A:Cross-references: GB:AB002446; GB:AB002098; NID:G7226185; PIDN:AAF41356.1; PID:G722618
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0950
C:Superfamily: fumarate reductase flavoprotein; 3-oxosteroid 1-dehydrogenase homology; 4

Query Match 70.0%; Score 35; DB 2; Length 587;
Best Local Similarity 70.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 EBYVPXGXDY 10
DB 366 EBYVPQGSBDY 375

RESULT 5
T48898

disease resistance protein RPP8 [validated] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 21-Jul-2000
C:Accession: T48898
R:McDowell, J.M.; Dhandaydham, M.; Long, T.A.; Aarts, M.G.; Goff, S.; Holub, E.B.; Dan
Plant Cell 10, 1861-1874, 1998
A:Title: Intragenic recombination and diversifying selection contribute to the evoluti
A:Reference number: Z24999; MUID:99030193; PMID:9811794
A:Accession: T48898
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-906 <MCD>
A:Cross-references: EMBL:AF089710; NID:G3928861; PIDN:AA683165.1; PID:G3928862
A:Experimental source: Landsberg erecta
C:Genetics:
A:Gene: RPP8
A:Introns: 293/1; 342/1
C:Function:
A:Description: promotes resistance to Peronospora parasitica

Query Match 70.0%; Score 35; DB 2; Length 906;
Best Local Similarity 60.0%; Pred. No. 40;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
OY 1 EBYVPXGXDY 10
DB 881 EBYVPQGSBDY 890

RESULT 6
T48899
disease resistance protein rpp8 [similarity] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 21-Jul-2000
C:Accession: T48899
R:McDowell, J.M.; Dhandaydham, M.; Long, T.A.; Aarts, M.G.; Goff, S.; Holub, E.B.; Dan
Plant Cell 10, 1861-1874, 1998
A:Title: Intragenic recombination and diversifying selection contribute to the evoluti
A:Reference number: Z24999; MUID:99030193; PMID:9811794
A:Accession: T48899
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-908 <MCD>
A:Cross-references: EMBL:AF089711; NID:G3901293; PIDN:AACT8631.1; PID:G3901294
A:Experimental source: Columbia
C:Genetics:
A:Gene: rpp8
A:Introns: 293/1; 342/1
C:Function:
A:Description: susceptible allele of a gene that promotes resistance to Peronospora pa

Query Match 70.0%; Score 35; DB 2; Length 908;
Best Local Similarity 60.0%; Pred. No. 40;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
OY 1 EBYVPXGXDY 10
DB 883 EBYVPQGSBDY 892

RESULT 7
A42452
V1 protein - tobacco yellow dwarf virus (strain Australia)
C:Species: tobacco yellow dwarf virus
C:Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C:Accession: A42452
R:Morrice, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
Virology 187, 633-642, 1992
A:Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yei
A:Reference number: A42452; MUID:92188538; PMID:11546458
A:Accession: A42452
A:Molecule type: DNA
A:Residues: 1-102 <MOR>

A:Cross-references: GB:M01103; NID:G335283; PIDN:AAA7947.1; PID:G335284

Query Match 68.0%; Score 34; DB 2; Length 102;

Best Local Similarity 60.0%; Pred. No. 6;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVFXGXDYS 11

DB 7 QVPSGINTS 16

RESULT 8

554619

hypothetical protein YOR013w - Yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein O2612; hypothetical protein YOL303.3

C:Species: Saccharomyces cerevisiae

C:Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002

C:Accession: S54619; S66879

R:De Haan, M.; Maarse, A.C.; Grivell, L.A.

A:Submitted to the EMBL data library, May 1995

A:Reference number: S54617

A:Accession: S54619

A:Molecule type: DNA

A:Residues: 1-156 <DB>

A:Cross-references: EMBL:X87331; NID:G1041652; PIDN:CAA60762.1; PID:G829123

R:De Haan, M.; Grivell, L.A.; Maarse, A.C.

A:Submitted to the Protein Sequence Database, July 1996

A:Reference number: S66877

A:Accession: S66879

A:Molecule type: DNA

A:Residues: 1-156 <DEW>

A:Cross-references: EMBL:Z74920; NID:G1420109; PIDN:CAA92201.1; PID:G1420111; MIPS:YOR01

A:Experimental source: strain 8286C

C:Genetics:

A:Cross-references: SGD:S0005539

A:Map position: 15R

C:Superfamily: hypothetical protein YOR013w

Query Match 68.0%; Score 34; DB 2; Length 156;

Best Local Similarity 66.7%; Pred. No. 9.6;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVFXGXDY 10

DB 50 EVVPLGMDY 58

RESULT 9

A96001

conserved hypothetical protein, homolog to osmotically inducible sensory protein SMC22-1

C:Species: Sinorhizobium meliloti

C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: A96001

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernat

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-fixing endo

A:Reference number: A95842; MUID:21396508; PMID:11481431

A:Accession: A96001

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-247 <KUR>

A:Cross-references: GB:AL591985; PIDN:CA49673.1; PID:G15141160; GSPDB:GN00167

A:Experimental source: strain 1021, megaplasmid pSymb

R:Gallibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,

hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yen, K.

A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: SMD21444

A:Genome: plasmid

Query Match 68.0%; Score 34; DB 2; Length 247;

Best Local Similarity 60.0%; Pred. No. 16;

Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BRVFXGXDYS 10

DB 48 EDVPRGADY 57

RESULT 10

A96546

unknown protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: A96546

R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K

ansen, N.F.; Hughes, B.; Hultzer, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luross, J.S.; Maiti, R.; Marziani

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: A96546

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-257 <STO>

A:Cross-references: GB:AB005173; NID:G11094688; PIDN:AAG39624.1; GSPDB:GN00141

C:Genetics:

A:Gene: P8A12.12

A:Map position: 1

Query Match 68.0%; Score 34; DB 2; Length 257;

Best Local Similarity 54.5%; Pred. No. 17;

Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BRVFXGXDYS 11

DB 217 BRUKAGADY 227

RESULT 11

F82491

ferrialdexophore reductase VCA0183 [imported] - Vibrio cholerae (strain N16961 serogro

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: F82491

R:Heidelberg, J.F.; Sisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gilm, M.L.; Dodson, R.J

Hardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers,

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: F82491

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-394 <HEI>

A:Cross-references: GB:AB004358; GB:AB003853; NID:G9657566; PIDN:AAF96096.1; GSPDB:GN0

A:Experimental source: serogroup O1, strain N16961, biotype El Tor

C:Genetics:

A:Gene: VCA0183

A:Map position: 2

C:Superfamily: flavohemoglobin; cytochrome-b5 reductase homology; globin homology

Query Match 68.0%; Score 34; DB 2; Length 394;

Best Local Similarity 66.7%; Pred. No. 27;

Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 EVVFXGXDX 10
 |||||
 Db 194 EVVPEGSXDY 202

RESULT 12

H87660
 peptidoglycan-binding protein, probable [imported] - Caulobacter crescentus
 C/Species: Caulobacter crescentus
 C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C/Accession: H87660
 R.Nieman, W.C.; Feildbyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gaitm, M.L.; Haft, D.H.; Kolon
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A/Title: Complete Genome Sequence of Caulobacter crescentus.
 A/Reference number: A87249; MUID:21173698; PMID:11259647
 A/Accession: H87660
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-433 <STO>
 A/Cross-references: GB:AE005673; NID:G13425020; PIDN:AAK25284.1; GSPDB:GN00148
 C/Genetics:
 A/Gene: CC3322

Query Match 68.0%; Score 34; DB 2; Length 433;
 Best Local Similarity 54.5%; Pred. No. 29;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 EEVVPXGXDXS 11
 |||||
 Db 266 EVILPFGFDYS 276

RESULT 13

B49132
 fac facets (faf) splice form 1 - fruit fly (Drosophila melanogaster)
 C/Species: Drosophila melanogaster
 C/Date: 19-Dec-1993 #sequence_revision 25-Apr-1997 #text_change 01-Dec-2000
 C/Accession: B49132; A49132
 R.Fischer-Vize, J.A.; Rubin, G.M.; Lehmann, R.
 Development 116, 985-1000, 1992
 A/Title: The fac facets gene is required for Drosophila eye and embryo development.
 A/Reference number: A49132; MUID:93202020; PMID:1295747
 A/Contents: isogenic BC
 A/Accession: B49132
 A/Status: preliminary
 A/Molecule type: nucleic acid
 A/Residues: 1-2747 <FIS>
 A/Cross-references: GB:I04959; NID:G157411; PIDN:AAF01345.1; PID:G6013474
 A/Note: sequence inconsistent with the nucleotide translation
 A/Note: sequence extracted from NCBI backbone (NCBIN:127836, NCBIIP:127839)
 A/Accession: A49132
 A/Status: preliminary
 A/Molecule type: nucleic acid
 A/Residues: 1-2704, 'VT', '2707', 'ANNV', <FIS>
 A/Cross-references: GB:I04958; NID:G157410; PIDN:AAF01346.1; PID:G6013475
 A/Note: sequence extracted from NCBI backbone (NCBIN:127836, NCBIIP:129008, NCBIIP:129029)
 C/Keywords: alternative splicing

Query Match 68.0%; Score 34; DB 2; Length 2747;
 Best Local Similarity 54.5%; Pred. No. 2.2e+02;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 EEVVPXGXDXS 11
 |||||
 Db 1394 EVVVPDGDGDS 1404

RESULT 14
 VKLJSI
 trans-regulatory splicing protein - simian immunodeficiency virus SIVcpz

N/Alternate names: anti-repression trans-activator; art protein; rev protein; trs prot.
 C/Species: simian immunodeficiency virus SIVcpz
 A/Note: host Pan troglodytes (chimpanzee)
 C/Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
 C/Accession: S09988
 R.Huet, T.; Cheynier, R.; Meyershan, A.; Roelants, G.; Wain-Hobson, S.
 Nature 345, 356-359, 1990

A/Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.
 A/Reference number: S09988; MUID:90259077; PMID:2188136
 A/Accession: S09988
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Cross-references: EMBL:X52154; NID:G58866; PIDN:CAA36405.1; PID:G763085
 C/Genetics:
 A/Gene: rev; trs; art
 A/Intons: 27/1
 C/Superfamily: AIDS trans-regulatory splicing protein
 C/Keywords: AIDS; immunodeficiency; splicing protein; transcription regulation

Query Match 66.0%; Score 33; DB 1; Length 124;
 Best Local Similarity 60.0%; Pred. No. 12;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 EVVFXGXDXS 11
 |||||
 Db 107 EVVPAAGNYS 116

RESULT 15

AG1272
 thiol peroxidases homolog lmo1583 [imported] - Listeria monocytogenes (strain EGD-e)
 C/Species: Listeria monocytogenes
 C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 17-May-2002
 C/Accession: AG1272
 R.Glaeser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Blocek
 -, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, J.
 Science 294, 849-852, 2001
 A/Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.;
 Ok, C.; Schluter, T.; Simoes, N.; Tixeret, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlan
 A/Title: Comparative genomics of Listeria species.
 A/Reference number: AB1077; MUID:21537279; PMID:11679669
 A/Accession: AG1272
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-165 <GLA>
 A/Cross-references: GB:NC_003210; PIDN:CAC99661.1; PID:G16411012; GSPDB:GN00177
 A/Experimental source: strain EGD-e
 C/Genetics:
 A/Gene: lmo1583
 C/Superfamily: thioredoxin peroxidase

Query Match 66.0%; Score 33; DB 2; Length 165;
 Best Local Similarity 66.7%; Pred. No. 17;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 EVVFXGXDX 10
 |||||
 Db 144 EVVPEGSXDH 152

Search completed: December 22, 2003, 17:45:01
 Job time: 10.0667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:43:51 / Search time 4.6 Seconds

(without alignments)
112.455 Million cell updates/sec

Title: US-09-909-164-45
Perfect score: 50
Sequence: 1 BEVVPXGXOXS 11

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_41*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	70.0	908	RBL4_ARATH	Q9FJX8 arabidopsis
2	35	70.0	908	RPP8_ARATH	Q6W4J9 arabidopsis
3	35	70.0	910	RPP8_ARATH	P59584 arabidopsis
4	35	70.0	1056	CARB_FUSN	Q87986 fusobacteri
5	34	68.0	102	YIK_TYDVA	P16193 tobacco yel
6	34	68.0	394	HMPA_VIBCH	Q9KMP3 vibrio chol
7	34	68.0	2778	FAF_DROME	P55824 drosophila
8	33	66.0	124	REV_SIVC2	P17280 chimpanzee
9	33	66.0	165	TPX_LISIN	Q92bc5 listeria in
10	33	66.0	165	TPX_LISIN	Q92bc5 listeria in
11	33	66.0	327	YK14_CAEEL	P34338 caenorhadi
12	33	66.0	421	ACDM_RAT	P38503 rattus norv
13	33	66.0	563	SVR_STRP3	Q6K5J2 streptococc
14	33	66.0	563	SVR_STRP6	Q6K5J2 streptococc
15	33	66.0	563	SVR_STRP6	Q6K5J2 streptococc
16	32	64.0	154	PLAS_ORYSA	P20423 oryza sativ
17	32	64.0	155	PLAS_HORYU	P08248 hordeum vul
18	32	64.0	168	PLAT_POPI	P11970 populus nig
19	32	64.0	196	YJAG_SALT1	Q91912 salmoneilla
20	32	64.0	427	TOLB_HABIN	P41672 haemophilus
21	32	64.0	1499	ALOC_HUMAN	O60312 homo sapien
22	32	64.0	3174	CHAC_HUMAN	Q96112 homo sapien
23	31	62.0	233	HIS9_THEMA	Q9WZ11 thermotoga
24	31	62.0	319	YHAI_CRIPIA	PI0941 cryptosporid
25	31	62.0	421	ACDM_HUMAN	P11310 homo sapien
26	31	62.0	421	ACDM_HUMAN	P11310 homo sapien
27	31	62.0	562	TRZM_PANAY	Q47861 panacea agq
28	30	60.0	97	PLAS_DAUCA	P20422 daucus caro
29	30	60.0	121	TYRK_HUMAN	Q9UHF0 homo sapien
30	30	60.0	165	YJ49_ARCFU	O28330 archaeglob
31	30	60.0	175	HRS3_RAT	Q04667 rattus norv
32	30	60.0	203	CTC_BACSV	P41919 bacillus su
33	30	60.0	231	ARAD_ECOLI	P08203 escherichia

34	30	60.0	231	1	ARAD_SALTY	P06190 salmoneilla
35	30	60.0	232	1	SCOA_HELPU	Q92163 helicobacte
36	30	60.0	232	1	SCOA_HELPU	P56006 helicobacte
37	30	60.0	259	1	OVUN_TYMT	P06308 lymnaea sta
38	30	60.0	421	1	AMP2_YEAST	P38174 saccharomyc
39	30	60.0	421	1	ECB2_HALEL	O52250 halomonas e
40	30	60.0	423	1	ECB1_HALEL	O92eu7 halomonas e
41	30	60.0	430	1	FOLC_BACSV	O05865 bacillus su
42	30	60.0	457	1	2185_HUMAN	O15231 homo sapien
43	30	60.0	469	1	LEFT_KUULA	P53998 kluyveromyc
44	30	60.0	478	1	AMP2_HUMAN	P50579 homo sapien
45	30	60.0	478	1	AMP2_MOUSE	O08663 mus musculu

ALIGNMENTS

RESULT 1
RBL4_ARATH STANDARD; PRT; 908 AA.
ID RBL4_ARATH
AC Q9FJX8
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable disease resistance RPP8-like protein 4.
GN RPP8L4 OR AT5G48620 OR K15N18.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=9087489; PubMed=9872454;
RA Nakamura Y., Sato S., Asamizu E., Kaneo T., Kotani H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
RT Sequence features of the regions of 1,013,767 bp covered by sixteen
RT physically assigned P1 and TAC clones.";
RL DNA Res. 5:297-308(1998).
[2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Hayashizaki Y., Shinozaki K.;
RT "Arabidopsis thaliana full-length cDNA.";
RT Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Potential disease resistance protein.
CC -1- DOMAIN: The LRR repeats probably act as specificity determinant of
CC pathogen recognition (by similarity).
CC -1- SIMILARITY: Belongs to the disease resistance NB-LRR family.
CC RPP8/HRT subfamily.
CC -1- SIMILARITY: Contains 3 leucine-rich (LRR) repeats.
CC -1- SIMILARITY: Contains 1 NB-ARC domain.
CC -1- DATABASE: NMR-NB-LRRS;
CC NOTS=Functional and comparative genomics of disease resistance gene
CC homologs;
CC WWW=http://nblrrs.ucdavis.edu".
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
DR EMBL: AB015468; BAB10695.1;
DR EMBL: AK117163; BAC41841.1;
DR InterPro: IPR000767; Disease_resist.
DR InterPro: IPR001611; LRR.

DR InterPro: IPR002182; NB-ARC.
 DR Pfam: PF00560; LRR; 2.
 DR Pfam: PF00931; NB-ARC; 1.
 DR PRINTS: PR00364; DISBASERISIT.
 KM Plant defense; ATP-binding; Repeat; Leucine-rich repeat.
 FT DOMAIN 10 45
 FT REPEAT 146 459
 FT REPEAT 575 599
 FT REPEAT 600 623
 FT REPEAT 842 867
 FT NP BIND 192 199
 SQ SEQUENCE 908 AA; 104448 MW; 3111991B17239693 CRC64;
 Query Match 70.0%; Score 35; DB 1; Length 908;
 Best Local Similarity 60.0%; Pred. No. 16;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 EEVVPXGXDY 10
 Db 883 EKLVPQSESDY 892

RESULT 2
 RPP8_ARATH STANDARD; PRT; 908 AA.
 ID RPP8_ARATH STANDARD; PRT; 908 AA.
 AC Q8W4J9; Q8W4J9; Q9ZSY3; Q9ZSY4;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Disease resistance protein RPP8 (Resistance to Peronospora parasitica
 protein 8).
 GN RPP8 OR HRT OR ATSG43470 OR MMF20.19.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, MUTANTS RPP8-1; RPP8-2 AND
 RP RPP8-3, AND VARIANTS.
 RC STRAIN=cv. Columbia, and cv. Landsberg erecta;
 RX MEDLINE=99030193; PubMed=9811794;
 RA McDowell J.M., Dhandaydham M., Long T.A., Aarss M.G.M., Goff S.,
 RA Holub E.B., Dangl J.L.;
 RT "Intergenic recombination and diversifying selection contribute to the
 RT evolution of downy mildew resistance at the RPP8 locus of
 RT Arabidopsis";
 RL Plant Cell 10:1861-1874 (1998).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS.
 RC STRAIN=cv. Di-17;
 RX MEDLINE=20271766; PubMed=10810142;
 RA Cooley M.B., Pachirana S., Wu H.-J., Kachroo P., Kleesig D.F.;
 RT "Members of the Arabidopsis HRT/RPP8 family of resistance genes confer
 RT resistance to both viral and oomycete pathogens";
 RL Plant Cell 12:663-676 (2000).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20181125; PubMed=10718197;
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
 RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
 RT clones";
 RL DNA Res. 7:31-63 (2000).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=cv. Columbia;
 RA Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 RT SSP consortium (Salk/Stanford/GRG)";
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

RN [5]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC STRAIN=cv. Columbia;
 RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
 RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
 RA Hayashizaki Y., Shinzaki K.;
 RT "Arabidopsis thaliana full-length cDNA";
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP INTERACTION WITH TIP.
 RX MEDLINE=20496823; PubMed=11041886;
 RA Ren T., Ou P., Morris T.J.;
 RT "HRT gene function requires interaction between a NAC protein and
 RT viral capsid protein to confer resistance to turnip crinkle virus";
 RL Plant Cell 12:1917-1926 (2000).
 CC -1- FUNCTION: Disease resistance protein. Resistance proteins guard
 CC the plant against pathogens that contain an appropriate avirulence
 CC protein via an indirect interaction with this avirulence protein.
 CC That triggers a defense system including the hypersensitive
 CC response, which restricts the pathogen growth. The interaction
 CC with TIP (TICV-interacting protein) may be essential for the
 CC recognition of the avirulence proteins, and the triggering of the
 CC defense response.
 CC -1- SUBUNIT: Interacts with the NAC protein TIP.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q8W4J9-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q8W4J9-2; Sequence=VSP_007171, VSP_007172;
 CC Note=Has been shown to exist only in cv. Columbia so far;
 CC -1- DOMAIN: The LRR repeats probably act as specificity determinant of
 CC pathogen recognition.
 CC -1- POLYMORPHISM: The strong polymorphisms present in cv. Di-17 and
 CC cv. Columbia are probably due to an unequal crossing-over between
 CC the highly related RPP8 and RPP8A genes present in cv. Landsberg
 CC erecta. Such variations probably modify the specificity of
 CC pathogen recognition.
 CC -1- MISCELLANEOUS: In cv. Columbia and cv. Landsberg erecta, RPP8
 CC specifically recognizes the Bmco5 avirulence protein from
 CC Peronospora parasitica, while it is not the case in cv. Di-17,
 CC where it confers resistance to Turnip Crinkle Virus upon
 CC recognition of the viral capsid protein.
 CC -1- SIMILARITY: Belongs to the disease resistance NB-LRR family.
 CC RPP8/HRT subfamily.
 CC -1- SIMILARITY: Contains 2 leucine-rich (LRR) repeats.
 CC -1- SIMILARITY: Contains 1 NB-ARC domain.
 CC -1- DATABASE: NAME=NTB-LRRS;
 CC NOTS=Functional and comparative genomes of disease resistance gene
 CC homologs;
 CC WWW=http://nblrrs.ucdavis.edu".
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AF089710; AAC83165.1; -
 CC EMBL: AF089711; AAC78631.1; -
 CC EMBL: AF234174; AAF36987.1; -
 CC EMBL: AB025638; BA97426.1; -
 CC EMBL: AY062514; AAL32592.1; -
 CC EMBL: AK118862; BAC43449.1; -
 CC EMBL: IPR000767; Disease_resist.
 CC InterPro: IPR001611; LRR.
 CC InterPro: IPR002182; NB-ARC.
 CC Pfam: PF00560; LRR; 2.
 CC PRINTS: PR00364; DISBASERISIT.
 KM Plant defense; ATP-binding; Repeat; Leucine-rich repeat;

KM Alternative splicing; Polymorphism.
 FT DOMAIN 10 45
 FT DOMAIN 146 459
 FT REPEAT 600 623
 FT REPEAT 842 867
 FT NP BIND 192 199
 FT VARSPLIC 294 308
 FT VARSPLIC 309 908
 FT VARIANT 27 31
 FT VARIANT 29 29
 FT VARIANT 85 87
 FT VARIANT 88 91
 FT VARIANT 100 100
 FT VARIANT 121 121
 FT VARIANT 129 129
 FT VARIANT 133 133
 FT VARIANT 138 138
 FT VARIANT 173 173
 FT VARIANT 177 177
 FT VARIANT 253 253
 FT VARIANT 257 257
 FT VARIANT 265 265
 FT VARIANT 267 267
 FT VARIANT 270 270
 FT VARIANT 283 283
 FT VARIANT 369 369
 FT VARIANT 387 387
 FT VARIANT 399 399
 FT VARIANT 426 426
 FT VARIANT 429 429
 FT VARIANT 436 446
 FT VARIANT 438 443
 FT VARIANT 450 450
 FT VARIANT 460 460
 FT VARIANT 480 486
 FT VARIANT 481 481
 FT VARIANT 485 486
 FT VARIANT 489 489
 FT VARIANT 514 514
 FT VARIANT 518 518
 FT VARIANT 519 519
 FT VARIANT 528 528
 FT VARIANT 537 537
 FT VARIANT 542 542
 FT VARIANT 543 543
 FT VARIANT 547 550
 FT VARIANT 549 549
 FT VARIANT 558 555
 FT VARIANT 558 560
 FT VARIANT 564 564
 FT VARIANT 584 584
 FT VARIANT 595 595
 FT VARIANT 609 612
 FT VARIANT 630 635
 FT VARIANT 632 636
 FT VARIANT 650 650
 FT VARIANT 653 654
 FT VARIANT 659 659
 FT VARIANT 660 660

LEUCINE-ZIPPER.
 NB-ARC.
 LRR 1.
 LRR 2.
 ATP (POTENTIAL).
 WKMLTSTSRNGVGH -> EILWYTHALFLINS (in isoform 2).
 /FTID-VSP 007171.
 Missing (in isoform 2).
 /FTID-VSP 007172.
 IDGLQ -> VDEQI (in cv. Landsberg erecta).
 G -> E (in cv. Di-17).
 SGK -> RGE (in cv. Di-17 and cv. Landsberg erecta).
 LKGV -> EKG (in cv. Landsberg erecta).
 C -> R (in cv. Landsberg erecta).
 E -> D (in cv. Landsberg erecta).
 F -> L (in cv. Di-17).
 Q -> QQ (in cv. Di-17).
 G -> V (in cv. Landsberg erecta).
 K -> T (in cv. Di-17).
 G -> C (in cv. Di-17).
 Y -> S (in cv. Di-17).
 R -> P (in cv. Di-17).
 A -> T (in cv. Di-17).
 R -> K (in cv. Landsberg erecta).
 V -> L (in cv. Di-17).
 V -> R (in cv. Di-17).
 A -> V (in cv. Di-17 and cv. Landsberg erecta).
 F -> S (in cv. Di-17 and cv. Landsberg erecta).
 W -> C (in cv. Di-17 and cv. Landsberg erecta).
 C -> R (in cv. Di-17 and cv. Landsberg erecta).
 N -> P (in cv. Di-17 and cv. Landsberg erecta).
 DSESTYSLEF -> YSKISAYDLFN (in cv. Landsberg erecta).
 BISTYS -> KITQOE (in cv. Di-17).
 A -> V (in cv. Landsberg erecta).
 E -> Q (in cv. Landsberg erecta).
 DNYLSMO -> NKTLRHH (in cv. Di-17).
 DN -> NR (in cv. Landsberg erecta).
 WQ -> SH (in cv. Landsberg erecta).
 Y -> N (in cv. Landsberg erecta).
 I -> K (in cv. Di-17 and cv. Landsberg erecta).
 C -> S (in cv. Di-17 and cv. Landsberg erecta).
 T -> I (in cv. Di-17).
 S -> R (in cv. Di-17).
 S -> R (in cv. Di-17).
 H -> Q (in cv. Di-17).
 I -> L (in cv. Landsberg erecta).
 KNKT -> RNNA (in cv. Di-17).
 K -> N (in cv. Landsberg erecta).
 PRPEDIW -> WDEDFG (in cv. Landsberg erecta).
 PRP -> SRPK (in cv. Di-17).
 Y -> F (in cv. Di-17).
 C -> Y (in cv. Di-17 and cv. Landsberg erecta).
 C -> S (in cv. Di-17 and cv. Landsberg erecta).
 YEAK -> FLAG (in cv. Landsberg erecta).
 NLRVDT -> DLSVHB (in cv. Di-17).
 RVDYB -> SVNNK (in cv. Landsberg erecta).
 Q -> E (in cv. Di-17).
 YL -> HI (in cv. Di-17).
 Missing (in cv. Di-17).
 M -> K (in cv. Landsberg erecta).

FT VARIANT 676 678
 FT VARIANT 678 679
 FT VARIANT 692 692
 FT VARIANT 697 697
 FT VARIANT 700 700

YLY -> FLF (in cv. Landsberg erecta).
 YG -> FR (in cv. Di-17).
 R -> H (in cv. Landsberg erecta).
 R -> Q (in cv. Di-17).
 A -> G (in cv. Di-17).

Query Match 70.0%; Score 35; DB 1; Length 908;
 Best Local Similarity 60.0%; Pred. No. 16;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPKXGY 10
 DB 883 EKLVPGEEDY 892

RESULT 3
 RPH ARATH
 ID RPH ARATH STANDARD; PRT; 910 AA.
 AC P59584;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DB Disease resistance protein RPH8A (RPH8 homolog A).
 GN RPH8A.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eubacteria; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A., AND FUNCTION.
 RX STRAIN=cv. Landsberg erecta;
 RX MEDLINE=9901093; PubMed=9811794;
 RA McDowell J.M., Dhandaydham M., Long T.A., Aarts M.G.M., Goff S.,
 RA Holub E.B., Dangl J.L.;
 RT "Intergenic recombination and diversifying selection contribute to the
 RT evolution of downy mildew resistance at the RPP8 locus of
 RT Arabidopsis.";
 RL Plant Cell 10:1861-1874 (1998).
 CC -1- FUNCTION: Disease resistance protein. Resistance proteins guard
 CC the plant against pathogens that contain an appropriate avirulence
 CC protein via an indirect interaction with this avirulence protein.
 CC That triggers a defense system including the hypersensitive
 CC response, which restricts the pathogen growth. In contrast to
 CC RPP8, it does not specifically recognize the Ercos avirulence
 CC protein from Pteronopora parasitica.
 CC MISCELLANEOUS: In cv. Columbia and cv. Di-17, this protein is not
 CC present due to an unequal crossing over between the RPP8 and RPH8A
 CC genes that creates a unique RPP8 gene.
 CC -1- SIMILARITY: Belongs to the disease resistance NB-LRR family.
 CC -1- RPP8/HRT subfamily.
 CC -1- SIMILARITY: Contains 2 leucine-rich (LRR) repeats.
 CC -1- SIMILARITY: Contains 1 NB-ARC domain.
 CC -1- DATABASE: NAME=NB-LRRs;
 CC NOTB=Functional and comparative genomics of disease resistance gene
 CC homologs;
 CC WWW=http://nhlhrs.ucdavis.edu".
 KW Plant defense; ATP-binding; Repeat; Leucine-rich repeat.
 FT DOMAIN 10 45
 FT DOMAIN 146 459
 FT REPEAT 602 625
 FT REPEAT 844 869
 FT NP BIND 192 199
 FT SEQUENCE 910 AA; 105263 MW; 5B1B9F5A19A12B CRC64;

Query Match 70.0%; Score 35; DB 1; Length 910;
 Best Local Similarity 60.0%; Pred. No. 16;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPKXGY 10
 DB 885 EKLVPGEEDY 894

```

RESULT 4
CARB_FUSNN STANDARD; PRT; 1058 AA.
ID CARB_FUSNN
AC O8RG86;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
phosphate synthetase ammonia chain).
OS CARB OR FN0422.
GN Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21866394; PubMed=11899109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Log T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grecklin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman B., Bernal A.,
RA Larsen N., D'Souza M., Malunas T., Pusch G., Haselkorn R.,
RA Forrester M., Kyriades N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
phosphate + L-glutamate + carbamoyl phosphate.
CC -1- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -1- PATHWAY: Arginine biosynthesis, first step.
CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain
promotes the hydrolysis of glutamine to ammonia, which is used by
the large (or ammonia) chain to synthesize carbamoyl phosphate (By
similarity).
CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR HAMAP: MF 01210; -; 1.
DR HAMAP: A010554; AAL94625.1; ALT_INIT.
DR InterPro: IPR006275; CarA_L_glu.
DR InterPro: IPR005483; CPase_L_D2.
DR InterPro: IPR005479; CPase_L_D2.
DR InterPro: IPR005480; CPase_L_D3.
DR InterPro: IPR005481; CPase_L_N.
DR InterPro: IPR004362; MCS like.
DR Pfam: PF00289; CPase_L_Chain; 2.
DR Pfam: PF02786; CPase_L_D2; 2.
DR Pfam: PF02787; CPase_L_D3; 1.
DR Pfam: PF02142; MCS; 1.
DR PRINTS: PR00098; CPASR.
DR TIGRFAMs: TIGR01369; CPASR1_lrg; 1.
DR PROSITE: PS00867; CPASR_2; 2.
DR PROSITE: PS00866; CPASR_1; 2.
DR Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1058
FT NP BIND 153 210 ATP (POTENTIAL).
FT NP BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).

```

```

FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1E39F CRC64;

Query Match
Best Local Similarity 60.0%; Score 34; DB 1; Length 102;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Cy 2 EYVPGXGDS 11
Db 190 EYVPGXGDS 199

RESULT 5
YLIK_TYDVA STANDARD; PRT; 102 AA.
ID YLIK_TYDVA
AC P1619;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Hypothetical 11.2 kDa protein.
GN V.
OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
OX NCBI_TaxID=31599;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92188538; PubMed=1546458;
RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.B.;
RT "The nucleotide sequence of the infectious cloned DNA component of
tobacco yellow dwarf virus reveals features of geminiviruses
infecting monocotyledonous plants."
RL Virology 187:633-642(1992).
CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: M81103; AAA47947.1; -.
DR PIR: A42452; A42452.
DR InterPro: IPR002621; Gemin1_mov.
DR Pfam: PF01708; Gemin1_mov; 1.
RM Hypothetical protein.
SQ SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF55B67 CRC64;

Query Match
Best Local Similarity 68.0%; Score 34; DB 1; Length 102;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Cy 2 EYVPGXGDS 11
Db 7 QVVPSSGINS 16

RESULT 6
HMPA_VIBCH STANDARD; PRT; 394 AA.
ID HMPA_VIBCH
AC Q9RWT3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Flavohemoprotein (Hemoglobin-like protein) (Flavohemoglobin).
GN HMP OR VC0183.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;

```



```

Query Match          68.0%; Score 34; DB 1; Length 2778;
Best Local Similarity 54.5%; Pred. No. 86;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY      1 EEVPEXGXDXYS 11
      ||| ||| |||
Db      1394 ETVDPGQDQDS 1404

RESULT 8
REV_SIVCZ STANDARD; PRT; 124 AA.
AC P17280;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE RRV protein (Anti-repression transactivator protein) (ART/TRS).
GN RRV.
OS Chimpazee immunodeficiency virus (SIV/cpz) (CIV).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_Taxid=11723;
RN [1]
RP SEQUENCE FROM N.A.
FX MEDLINE=90259077; PubMed=2188136;
RA Hueb T., Cheynier R., Meyerhans A., Roelants G., Wain-Hobson S.;
RT "Genetic organization of a chimpanzee lentivirus related to HIV-1.";
RL Nature 345:356-359(1990).
CC -1- FUNCTION: RRV APPEARS TO ACT POST-TRANSCRIPTIONALLY TO RELIEVE
CC NEGATIVE REGRESSION OF GAG AND ENV PRODUCTION.
CC -1- SUBCELLULAR LOCATION: Nuclear; accumulates in the nucleol.
CC -1- PMW: PHOSPHOPROTEIN WHOSE STATE OF PHOSPHORYLATION IS MEDIATED
CC BY A SPECIFIC SRINE KINASE ACTIVITY PRESENT IN THE NUCLEUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52154; CAA36405.1; -.
DR PIR; G09988; VKLJST.
DR HIV; X52154; REVSCP2.
DR InterPro; IPR00625; REV_Protein.
DR Pfam; PF00424; REV; 1.
KW Transcription regulation; AIDS; Phosphorylation; Nuclear protein.
SQ SEQUENCE 124 AA; 13701 MW; F5877D1BDF65A7B2 CRC64;

Query Match          66.0%; Score 33; DB 1; Length 124;
Best Local Similarity 60.0%; Pred. No. 5.4;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0.

OY      2 EEVPEXGXDXYS 11
      ||| ||| |||
Db      107 ETVDPGQDQDS 116

RESULT 9
TPX_LISIN STANDARD; PRT; 165 AA.
ID _TPX_LISIN
AC Q92BC5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable thiol peroxidase (EC 1.11.1.-).
GN TPX OR LIN1625.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
NCBI_Taxid=1642;
RN [1]
RP SEQUENCE FROM N.A.

```

RA STRAIN-CLIP 11262 / Serovar 6a;
 RX MEDLINE=21537279; PubMed=11679669;
 RA Glaeser P., Frangoul L., Buchrieser C., Rusniok C., Amend A.,
 RA Baquero F., Berche P., Bloeker H., Brandt P., Charraborty T.,
 RA Chablit A., Chetouani F., Couve E., de Daruvar A., Deloux P.,
 RA Domnan K.-D., Fshil H., Garcia-del Portillo F., Dussauget O.,
 RA Gautier L., Geobel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,
 RA Jones L.-M., Kaerst U., Kreft J., Kunz M., Kuntz F., Kurupkat G.,
 RA Madeno E., Maltournam A., Mata Vicente J., Ng B., Nedjati H.,
 RA Nordsted G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schueter T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.,
 RA "Comparative genomes of *Listeria* species."
 RL Science 294:849-852(2001)
 CC -1- FUNCTION: Has antioxidant activity. Could remove peroxidases or
 H(2)O(2) (By similarity).
 CC -1- SIMILARITY: Belongs to the AhpC/Tsa family. Tpx subfamily.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

CC EMBL, AL596163; CAC96856.1; --
 DR PIR, AH1635; AH1635.
 DR L161181; L161181; --
 DR HAMAP, MF_00269; -- 1.
 DR InterPro, IPR000866; AhpC-TSA.
 DR InterPro, IPR002065; Tpx.
 DR Pfam, PF00578; AhpC-TSA; 1.
 DR PROSITE, PS01265; Tpx; 1.
 KW Antioxidant; Oxidoreductase; Peroxidase; Complete proteome.
 SQ SEQUENCE 165 AA; 18162 MW; 77705B7CDB8B6F4D CRC64;

Query Match 66.0%; Score 33; DB 1; Length 165;
 Best Local Similarity 66.7%; Pred. No. 7.3;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 EVVPGXGDY 10
 |||||
 DB 144 EVVPGSDH 152

RESULT 10
 ID TPX_LISMO STANDARD; PRT; 165 AA.
 AC Q816U8;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable thiol peroxidase (EC 1.11.1.-).
 GN TPX OR LMO1583.
 OS *Listeria monocytogenes*.
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 OC NCBI_TaxID=1639;
 RX MEDLINE=21537279; PubMed=11679669;
 RA Glaeser P., Frangoul L., Buchrieser C., Rusniok C., Amend A.,
 RA Baquero F., Berche P., Bloeker H., Brandt P., Charraborty T.,
 RA Chablit A., Chetouani F., Couve E., de Daruvar A., Deloux P.,
 RA Domnan K.-D., Fshil H., Garcia-del Portillo F., Dussauget O.,
 RA Gautier L., Geobel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,
 RA Jones L.-M., Kaerst U., Kreft J., Kunz M., Kuntz F., Kurupkat G.,
 RA Madeno E., Maltournam A., Mata Vicente J., Ng B., Nedjati H.,
 RA Nordsted G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schueter T., Simoes N., Tierrez A.,

RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.,
 RL Science 294:849-852(2001)
 CC -1- FUNCTION: Has antioxidant activity. Could remove peroxidases or
 H(2)O(2) (By similarity).
 CC -1- SIMILARITY: Belongs to the AhpC/Tsa family. Tpx subfamily.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

CC EMBL, AL591979; CAC99661.1; --
 DR PIR, AG1272; AG1272.
 DR L161181; L161181; --
 DR HAMAP, MF_00269; -- 1.
 DR InterPro, IPR000866; AhpC-TSA.
 DR InterPro, IPR002065; Tpx.
 DR Pfam, PF00578; AhpC-TSA; 1.
 DR PROSITE, PS01265; Tpx; 1.
 KW Antioxidant; Oxidoreductase; Peroxidase; Complete proteome.
 SQ SEQUENCE 165 AA; 18133 MW; 77705B7C046D424D CRC64;

OY 2 EVVPGXGDY 10
 |||||
 DB 144 EVVPGSDH 152

RESULT 11
 ID YK14_CABEL STANDARD; PRT; 327 AA.
 AC P34338;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein C15H7.4 in chromosome III.
 GN C15H7.4.
 OS *Caenorhabditis elegans*.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OC NCBI_TaxID=6239;
 RX MEDLINE=7906398;
 RA Wilson R., Atencough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Hawkins T., Hillier L., Jier M.,
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
 RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer B., Steden R.,
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
 RA Woldman P.,
 RL "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*."
 RL Nature 368:32-38 (1994).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

DR InterPro; IPR005148; N.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF03485; N-Arg; 1.
 DR Pfam; PF00750; tRNA-synt_1d; 1.
 DR PRINTS; PR01038; TRNASYNTHARG.
 DR TIGRFAMs; TIGR00456; args; 1.
 DR PROSITE; PS00178; AA tRNA LIGASE I; FALSE NEG.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 FT SITE 121 131 "HIGH" REGION.
 SQ SEQUENCE 563 AA; 63134 MW; B0B7DEC31A9DCF63 CRC64;
 Query Match 66.0%; Score 33; DB 1; Length 563;
 Best Local Similarity 45.5%; Pred. No. 26;
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 BEVVPXGDYS 11
 Db 94 BOVITAGSDYA 104
 RESULT 14
 SYR_STRPB STANDARD; PRT; 563 AA.
 AC Q8NZ22;
 DT 28-FEB-2003 (Rel. 41, Last Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ARGRS).
 GN ARGS OR SPY18 2183.
 OS Streptococcus pyogenes (serotype M18).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 CX NCBI_TaxID=186103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MGAS8232 / Serotype M18;
 RX MEDLINE=21927593; PubMed=11917108;
 RA Smoot J.C., Barbhan K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
 RA Sylva G.L., Studevant D.B., Ricklets S.M., Porcella S.F.,
 RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
 RA Kaur V., Daly J.A., Veasy L.G., Musser J.M.;
 RT "Genome sequence and comparative microarray analysis of serotype M18
 group A Streptococcus strains associated with acute rheumatic fever
 outbreaks.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
 CC -1- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
 CC diphosphate + L-arginyl-tRNA(Arg).
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB010119; AAL98627.1; -
 DR HAMAP; MF_00123; -; 1.
 DR InterPro; IPR001278; Arg_tRNA-synt_1c.
 DR InterPro; IPR005148; N.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF03485; N-Arg; 1.
 DR Pfam; PF00750; tRNA-synt_1d; 1.
 DR PRINTS; PR01038; TRNASYNTHARG.
 DR TIGRFAMs; TIGR00456; args; 1.
 DR PROSITE; PS00178; AA tRNA LIGASE I; FALSE NEG.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 FT SITE 121 131 "HIGH" REGION.
 SQ SEQUENCE 563 AA; 63134 MW; B0B7DEC31A9DCF63 CRC64;

SQ SEQUENCE 563 AA; 63134 MW; FECA176A68D85B CRC64;
 Query Match 66.0%; Score 33; DB 1; Length 563;
 Best Local Similarity 45.5%; Pred. No. 26;
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 BEVVPXGDYS 11
 Db 94 BOVITAGSDYA 104
 RESULT 15
 SYR_STRPY STANDARD; PRT; 563 AA.
 AC Q99XLS;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ARGRS).
 GN ARGS OR SPY2151.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 CX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
 RX MEDLINE=21192684; PubMed=11296296;
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
 RA Primeaux C., Seaton S., Suvorov A.N., Kenon S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
 CC -1- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
 CC diphosphate + L-arginyl-tRNA(Arg).
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB006633; AAK34788.1; -
 DR HSPR; Q05506; 18S2.
 DR HAMAP; MF_00123; -; 1.
 DR InterPro; IPR001278; Arg_tRNA-synt_1c.
 DR InterPro; IPR005148; N.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF03485; N-Arg; 1.
 DR Pfam; PF00750; tRNA-synt_1d; 1.
 DR PRINTS; PR01038; TRNASYNTHARG.
 DR TIGRFAMs; TIGR00456; args; 1.
 DR PROSITE; PS00178; AA tRNA LIGASE I; FALSE NEG.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 FT SITE 121 131 "HIGH" REGION.
 SQ SEQUENCE 563 AA; 63120 MW; B0F2CAC28D03B613 CRC64;
 Query Match 66.0%; Score 33; DB 1; Length 563;
 Best Local Similarity 45.5%; Pred. No. 26;
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 BEVVPXGDYS 11
 Db 94 BOVITAGSDYA 104

Tue Dec 23 09:51:20 2003

us-09-909-164-45.rsp

Page 10

Search completed: December 22, 2003, 17:42:29
Job time : 4.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:27:26 ; Search time 25.2 Seconds

(without alignments)
112.642 Million cell updates/sec

Title: US-09-909-164-45
Perfect score: 50
Sequence: 1 EEVVPXGXDYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mbc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeophages:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	72.0	298	2	052367 rhizobium t
2	36	72.0	319	16	08BSY7 Oesv7 oceanobacti
3	36	72.0	363	17	030260 O30260 archaeoglob
4	36	72.0	595	10	094637 094637 androgaphi
5	36	72.0	1044	16	08DIH0 Qdih0 synecococc
6	35	70.0	78	6	09XST4 Q9xst4 canis fami
7	35	70.0	143	5	09VSY8 Q9vys8 dirosophila
8	35	70.0	175	6	08I033 O8i033 bos taurus
9	35	70.0	215	6	08I031 O8i031 bos taurus
10	35	70.0	217	4	000404 O00404 homo sapien
11	35	70.0	290	16	08U770 Q8u770 agrobacteri
12	35	70.0	297	6	08HX19 Q8hx19 bos taurus
13	35	70.0	299	4	09UEB9 Q9uee9 homo sapien
14	35	70.0	587	16	09JZP8 Q9jzp8 neisseria m
15	35	70.0	692	5	09VI90 Q9vi90 dirosophila
16	35	70.0	906	10	09ZSY4 Q9zsy4 arabidopsis

17	35	70.0	908	10	Q8W4J9 Q8w4j9 arabidopsis
18	35	70.0	908	10	Q9ZSY3 Q9zsy3 arabidopsis
19	35	70.0	908	10	Q9FJX8 Q9fjx8 arabidopsis
20	35	70.0	909	10	Q9MSA1 Q9msa1 arabidopsis
21	34	68.0	156	3	Q12479 Q12479 saccharomyc
22	34	68.0	175	10	Q8VY88 Q8vy88 arabidopsis
23	34	68.0	247	16	Q92066 Q92066 rhizobium m
24	34	68.0	257	10	Q9C6J0 Q9c6j0 arabidopsis
25	34	68.0	283	16	Q8NR16 Q8nr16 corynebacte
26	34	68.0	433	16	Q9A382 Q9a382 caulobacter
27	34	68.0	1442	17	Q96VH5 Q96vh5 sulfolobus
28	33	66.0	143	17	Q8TX62 Q8tx62 methanopyru
29	33	66.0	196	16	Q8ZAQ9 Q8zaq9 yersinia pe
30	33	66.0	210	10	Q6S890 Q6s890 cyclostell
31	33	66.0	222	16	Q8E5L9 Q8e5l9 streptococ
32	33	66.0	222	16	Q8DZM9 Q8dzm9 streptococ
33	33	66.0	225	10	Q40129 Q40129 lycopersico
34	33	66.0	253	16	Q8XP88 Q8xpa8 clostridium
35	33	66.0	283	16	Q8FPY9 Q8fpy9 corynebacte
36	33	66.0	312	11	Q9D876 Q9d876 mus musculu
37	33	66.0	368	2	Q8K1A3 Q8k1a3 rhizobium e
38	33	66.0	440	17	Q9YF13 Q9yf13 aeropyrum p
39	33	66.0	471	11	Q8R126 Q8r126 mus musculu
40	33	66.0	484	11	Q8VD18 Q8vd18 mus musculu
41	33	66.0	484	11	Q8BXT4 Q8bxt4 mus musculu
42	33	66.0	484	11	Q8BK35 Q8bk35 mus musculu
43	33	66.0	517	16	Q8XZ15 Q8xz15 ralsconia s
44	33	66.0	563	13	Q9DDJ4 Q9ddj4 halichoeres
45	33	66.0	678	12	Q9E1X6 Q9e1x6 cercopithec

ALIGNMENTS

RESULT 1	052367	PRELIMINARY;	PRT;	298 AA.
ID	052367			
AC	052367			
DT	01-JUN-1998	(TRENBLREL. 06, Created)		
DT	01-JUN-1998	(TRENBLREL. 06, Last sequence update)		
DT	01-MAR-2003	(TRENBLREL. 23, Last annotation update)		
DE	Ar1-alcohol dehydrogenase homolog (Fragment).			
GN	XV181.			
OS	Rhizobium tropici.			
OC	Plasmid pRitCEN298.			
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;			
OC	Rhizobiaceae; Rhizobium.			
OX	NCBI_TaxID=398;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CEN298;			
RA	Rosenbluth M., Hynes M.F., Martinez-Romero E.;			
RL	Submitted (DBC-1997) to the EMBL/Genbank/DBJ databases.			
CC	-1 COPFACTOR: ZINC (BY SIMILARITY).			
CC	-1 SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE			
CC	FAMILY.			
DR	EMBL; AF036920; AAC04779.1; -.			
DR	InterPro; IPR002328; ADH_zinc.			
DR	InterPro; IPR002085; Adh_zn family.			
DR	InterPro; IPR000205; NAD_binding.			
DR	Pfam; PF00107; adh_zinc.1.			
DR	PROSITE; PS00059; ADH_ZINC.1.			
KW	Metal-binding; Oxidoreductase; zinc; Plasmid.			
FT	NON_TER 298			
SC	SEQUENCE 298 AA; 31092 MW; 49B2F8117C33AB87 CRC64;			
Query Match	72.0%; Score 36; DB 2; Length 298;			
Best Local Similarity	50.0%; Pred. No. 15;			
Matches	5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;			
QY	2 EEVVPXGXDYS 11			
DB	250 E1EFGADFS 259			

RESULT 2

Q8ESV7 PRELIMINARY; PRT; 319 AA.
 AC Q8ESV7;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Hypothetical conserved protein.
 GN O80509.
 OS Oceanobacillus theyensis.
 OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.
 OX NCBI_TaxID=182710;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
 RX MEDLINE=2220767; Pubmed=12235376;
 RA Takami H., Takaki Y., Uchiyama I.,
 RT "Genome sequence of Oceanobacillus theyensis isolated from the Iheya
 RT Ridge and its unexpected adaptive capabilities to extreme
 RT environments."
 RL Nucleic Acids Res. 30:3927-3935 (2002).
 DR EMBL: AP004594; BAC12465.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 319 AA; 35617 MW; 3BDAB4BF1E79E37 CRC64;

Query Match 72.0%; Score 36; DB 16; Length 319;
 Best Local Similarity 60.0%; Pred. No. 16;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVVVPXGXDX 10
 Db 189 EQLVPHGIDY 198

RESULT 3

O30260 PRELIMINARY; PRT; 363 AA.
 AC O30260;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 20, Last annotation update)
 DE Hypothetical protein AF2411.
 GN AF2411.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
 OC Archaeoglobaceae; Archaeoglobus.
 OX NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; Pubmed=9389475;
 RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.B.,
 RA Richardson K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirschner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Beich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.P., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Arlisch P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.,
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon Archaeoglobus fulgidus."
 RL Nature 390:364-370 (1997).
 DR EMBL: AE001109; AAB91255.1; -
 DR TIGR: AF2411; -
 DR InterPro: IPR002103; Bac_luciferase.
 DR Pfam: PF00296; bac_luciferase; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 363 AA; 41736 MW; 0E976AE788F4803 CRC64;

Query Match 72.0%; Score 36; DB 17; Length 363;
 Best Local Similarity 54.5%; Pred. No. 18;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 EVVVPXGXDXS 11
 Db 120 ENIVPGIDFS 130

RESULT 4

O946J7 PRELIMINARY; PRT; 595 AA.
 AC O946J7;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE 3-hydroxy-3-methylglutaryl-coenzyme A reductase (EC 1.1.1.34).
 GN HMGR1.
 OS Andropogon paniculata.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Aseridae; Lamiales; Acanthaceae; Acanthaceae; Ruelliales;
 OC Andropogonaceae; Andropogonaceae.
 OX NCBI_TaxID=175694;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Krishnan S., Banerjee N.S.,
 RT "3-hydroxy-3-methylglutaryl coenzyme A reductase gene from
 RT Andropogon paniculata."
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF389879; AL28015.2; -
 KW Oxidoreductase.
 SQ SEQUENCE 595 AA; 63268 MW; 19A3EA572F67AB2E CRC64;

Query Match 72.0%; Score 36; DB 10; Length 595;
 Best Local Similarity 70.0%; Pred. No. 33;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 EVVVPXGXDXS 11
 Db 15 EVAPPGHDXS 24

RESULT 5

O8DIH0 PRELIMINARY; PRT; 1044 AA.
 AC O8DIH0;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Multidrug efflux transporter.
 GN TLL1618.
 OS Synecchococcus elongatus (Thermosynechococcus elongatus).
 OC Bacteria; Cyanobacteria; Chroococcales; Synecchococcus.
 OX NCBI_TaxID=32046;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP-1;
 RX MEDLINE=2225144; Pubmed=12240834;
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 RA Wakemura A., Iriuchi M., Kawashima K., Kimura T., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
 RA Shimizu S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the thermophilic cyanobacterium
 RT Thermosynechococcus elongatus BP-1."
 RL DNA Res. 9:123-130 (2002).
 DR EMBL: AP005374; BAC09170.1; -
 KW Complete proteome.
 SQ SEQUENCE 1044 AA; 113205 MW; 00B9C33F0F636D2F CRC64;

Query Match 72.0%; Score 36; DB 16; Length 1044;
 Best Local Similarity 63.6%; Pred. No. 62;

Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXYS 11
DB 843 BEVLPNGIGYS 853

RESULT 6
QXST4 PRELIMINARY; PRT; 78 AA.

AC QXST4; 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
GN p97 homologous protein (Fragment).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thyroid;
RX MEDLINE=20422104; PubMed=10964405;
RA Pichon B., Mercan D., Pouillon V., Christophe-Hobertus C.,
Christophe D.,
RT "A method for the large-scale cloning of nuclear proteins and nuclear
RT targeting sequences on a functional basis."
RL Anal. Biochem. 284:231-239 (2000).
DR EMBL; AJ388531; CAB46833.1; -.
FT NON_TER 78
SQ SEQUENCE 78 AA; 8895 MW; B62486313555FBAB CRC64;

Query Match 70.0%; Score 35; DB 6; Length 78;
Best Local Similarity 54.5%; Pred. No. 5.3;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXYS 11
DB 16 EDVVPSCGYS 26

RESULT 7
QVSY8 PRELIMINARY; PRT; 143 AA.

AC QVSY8; 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CG3911 protein.
GN CG3911.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Ramanakides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Vandeil M.D., Zhang L.X.,
Brandon R.C., Rogers J.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
Adair J.F., Agayari A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,
Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brodeur P., Brotter P.,
Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablo B., Delcher A., Deng Z., Dew I., Dietz S.M.,

RA Dodson K., Doud L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
R Foster C., Gabriellian A.S., Gary N.S., Gelbart W.M., Glaeser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jallil M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulip D., Lai Z.,
RA Laske P., Lei Y., Levinsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Moadary C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunderson R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Svirskas R., Tector C., Turner R., Venter B., Wang A.H., Wang X.,
RA Wang Z.-Y., Wastarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
DR EMBL; AF003552; AAF50270.1; -.
DR FlyBase; FBgn0035922; CG3911.
SQ SEQUENCE 143 AA; 16471 MW; 1DC346DC22C02NA2 CRC64;

Query Match 70.0%; Score 35; DB 5; Length 143;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGXGD 9
DB 135 BEVLPAGHD 143

RESULT 8
Q81033 PRELIMINARY; PRT; 175 AA.

AC Q81033; 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Human-type bcl2 protein.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Jersey;
RA Iwashita S., Itoh T.,
RT "A LINE-mediated gene diversity."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB081003; BAC11952.1; -.
SQ SEQUENCE 175 AA; 19529 MW; C872803CB96393B6 CRC64;

Query Match 70.0%; Score 35; DB 6; Length 175;
Best Local Similarity 54.5%; Pred. No. 13;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXYS 11
DB 16 EDVVPSCGYS 26

RESULT 9
Q81031 PRELIMINARY; PRT; 215 AA.

AC Q81031; 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Human-type bcl2 protein.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Jersey;
RA Iwashita S., Itoh T.,
RT "A LINE-mediated gene diversity."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB081003; BAC11952.1; -.
SQ SEQUENCE 175 AA; 19529 MW; C872803CB96393B6 CRC64;

```

DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE H-type bent protein (Fragment).
GN H-TYPE BCNT.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RPZD #750.
RA Iwashita S., Itoh T., Sezaki M., Oshima K., Hashimoto K.,
RA Kitagawa Y., Takahashi T., Masui T., Hashimoto K.,
RA "A LINE-mediated gene diversity."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB081095; BAC15593.1; -.
FT NON TER 215
SQ SEQUENCE 215 AA; 23822 MW; 7287C8B98ED05C8 CRC64;

Query Match 70.0%; Score 35; DB 6; Length 215;
Best Local Similarity 54.5%; Pred. NO. 17;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
DB 16 EDVVPSCGERS 26

RESULT 10
ID 000404 PRELIMINARY; PRT; 217 AA.
AC 000404;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE p97 homologous protein.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=97160586; PubMed=9006920;
RA Nobukuni T., Kobayashi M., Oomori A., Ichinose S., Iwanaga T.,
RA Takahashi I., Hashimoto K., Hattori S., Kaibuchi K., Miyata Y.,
RA "An Alu-linked repetitive sequence corresponding to 280 amino acids is
RA expressed in a novel bovine protein, but not in its human homologue."
RL U. Biol. Chem. 272:2801-2807(1997).
DR EMBL; D85939; BAA20069.1; -.
SQ SEQUENCE 217 AA; 24061 MW; E404BA2B35497828 CRC64;

Query Match 70.0%; Score 35; DB 4; Length 217;
Best Local Similarity 54.5%; Pred. NO. 17;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGXDYS 11
DB 16 EDVVPSCGERS 26

RESULT 11
ID 08U7J0 PRELIMINARY; PRT; 230 AA.
AC 08U7J0;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE 6-O-methylguanine-DNA methyltransferase.
GN ADA OR ATU4459 OR AGR_L_818.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).

```

```

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OC NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Seubald J.C., Kaul R., Monks D.B., Kitajima J.P., Woo L.,
RA Okura V.K., Zhou Y., Chen L., Wood G.R., Almeida N.F. Jr., Moo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D. Sr.,
RA Chapman P., Clendenning J., Deacher G., Gilllet W., Grant C.,
RA Rutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chunley P., Tingey S.V., Tomb J.-P., Gordon M.P., Olson M.V.,
RA Neeter B.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Akenazi M., Halling C., Mullin L.,
RA Hounmel K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappe C., Markelz B.,
RA Piansgan C., Crowell C., Gerson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328 (2001).
DR EMBL; AB009374; AAL45253.1; -.
DR EMBL; AB008240; AAK89882.1; -.
DR InterPro; IPR000005; HTMAAC.
DR InterPro; IPR001497; Methyltransf_1.
DR Pfam; PF00165; HTH_ARAC; 2.
DR Pfam; PF01035; Methyltransf_1; 1.
DR SMART; SM00342; HTH_ARAC; 1.
DR TIGRfam; TIGR00589; ogf_1.
DR PROSITE; PS01124; HTH_ARAC_FAMILY_2; 1.
KW Methyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 290 AA; 31587 MW; B626592BF51977F CRC64;

Query Match 70.0%; Score 35; DB 16; Length 290;
Best Local Similarity 50.0%; Pred. NO. 24;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGXDY 10
DB 9 EDTTPGSDY 18

RESULT 12
ID 08HX9 PRELIMINARY; PRT; 297 AA.
AC 08HX9;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Human-type Bcmt.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Iwashita S., Itoh T.;
RA "A LINE-mediated gene diversity."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB081004; BAC11953.1; -.
SQ SEQUENCE 297 AA; 33354 MW; D4A944BC8740373C CRC64;

```

```

Query Match      70.0%; Score 35; DB 6; Length 297;
Best Local Similarity 54.5%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      1 EEVVPXGXDYS 11
        :|:|:|:|:|
        16 EDVVPSCGGEYS 26

RESULT 13
Q9UEB9      PRELIMINARY; PRT; 299 AA.
ID Q9UEB9
AC Q9UEB9
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE BCLT protein (CRANIOFACIAL development protein 1).
GN BCLT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=98267221; PubMed=9602175;
RA Takahashi I., Nobuhumi T., Ohmori H., Kobayashi M., Tanaka S.,
RA Ohshima K., Okada N., Masui T., Hashimoto K., Iwashita S.,
RT "Existence of a bovine LINB repetitive insert that appears in the cDNA
RL of bovine protein BCLT in ruminant, but not in human, genomes."
RN Gene 211:387-394(1998).
[2]
RN SEQUENCE FROM N.A.
RP
RC TISSUE=Placenta;
RA Straube R.;
RL Submitted (NCV-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB009285; BAA11867.1; -
DR EMBL; BC000991; AAH00991.1; -
DR Genew; HGNC:1873; CPDPI.
SQ SEQUENCE 299 AA; 33593 MW; FA9B928B669451A CRC64;

Query Match      70.0%; Score 35; DB 4; Length 299;
Best Local Similarity 54.5%; Pred. No. 25;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      1 EEVVPXGXDYS 11
        :|:|:|:|:|
        16 EDVVPSCGGEYS 26

Db      16 EDVVPSCGGEYS 26

RESULT 14
Q9JZP8      PRELIMINARY; PRT; 587 AA.
ID Q9JZP8
AC Q9JZP8
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Succinate dehydrogenase, flavoprotein subunit.
GN NMB00950.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OC NCBI_TaxID=491;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tetteijn H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.B.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwin M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Helt D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Maon T., Ciesko A., Parksey D.S., Blair E., Ciccione H., Clark E.B.,
RA Cotton M.D., Uterback T.R., Khouri H., Qin H., Vamathavan J.,

```

```

RA Gill J., Scarlato V., Maignani V., Piazza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon R.R., Rappoli R., Venter J.C.,
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
CC -1-COFACITOR: FAD (BY SIMILARITY).
DR EMBL; AB002446; AAF41356.1; -.
DR HGSP; P00363; IKF6.
DR TIGR; NMB0950; IKF6.
DR InterPro; IPR003953; FAD_bind2.
DR InterPro; IPR001327; FAD_Pyr_redox.
DR InterPro; IPR003952; PRD/SDH_FAD.
DR InterPro; IPR001100; Pyr_redox.
DR InterPro; IPR004112; Succ_DH_flav_C.
DR Pfam; PF00890; FAD_binding_2; 1.
DR Pfam; PF02910; succ_DH_flav_C; 1.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00411; PNDPRDASRI.
DR PROSITE; PS00504; PRD_SDH_FAD_BINDING; 1.
RV FAD; Flavoprotein; Oxidoreductase; Complete proteome.
SQ SEQUENCE 587 AA; 64502 MW; 9581701508069003 CRC64;

Query Match      70.0%; Score 35; DB 16; Length 587;
Best Local Similarity 70.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 EEVVPXGXDY 10
        :|:|:|:|:|
        366 EVVVPQGEYD 375

Db      366 EVVVPQGEYD 375

RESULT 15
Q9VT90      PRELIMINARY; PRT; 692 AA.
ID Q9VT90
AC Q9VT90
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE CG10040 protein (Roughened eye).
GN RN OR CG10040 OR CG14600 OR CG14601 OR CG14603 OR CG12466.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN (1)
RN SEQUENCE FROM N.A.
RP STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.B., Li P.W., Hopkins R.A., Galle R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter B.G., Holt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Baer A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintinas S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Buttle K.C., Buam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kemisun J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Liao X., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mervinov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,

```

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun B.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers B.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).

[2]

RP SEQUENCE FROM N.A.
 RA St Pierre S.B., Galindo M.I., Couso J.P., Thor S.;
 RT "Control of *Drosophila* imaginal disc development by rotund and
 RT roughened eye: differentially expressed transcripts of the same gene
 RT encoding distinct zinc finger proteins.";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AE003672; AAF54035.1; -;
 DR EMBL; AF395904; AAL59598.1; -;
 DR HSSP; P08153; 1ZPD.

DR FlyBase; FBgn0003263; zn.
 DR InterPro; IPR007087; Znf_C2H2.

DR Pfam; PF00096; ZF-C2H2; 5.

DR ProDom; PD000003; Znf_C2H2; 1.

DR SMART; SM00355; Znf_C2H2; 5.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.

DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 5.

KW Metal-binding; Zinc; Zinc-finger.
 SQ SEQUENCE 692 AA; 75774 MW; BD010502BB650428 CRC64;

Query Match 70.0%; Score 35; DB 5; Length 692;
 Best Local Similarity 60.0%; Pred. No. 65;

Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 2 EVVFXGXDYS 11
 | : | | | |

Db 26 RPPAGGDYS 35

Search completed: December 22, 2003, 17:51:33
 Job time : 26.2667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:41:00 ; Search time 32.4667 Seconds
(without alignments)
53.778 Million cell updates/sec

Title: US-09-909-164-46
Perfect score: 50
Sequence: 1 ERYVFXGXQDYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1980.DAT.*
 - 2: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1981.DAT.*
 - 3: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1982.DAT.*
 - 4: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1983.DAT.*
 - 5: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1984.DAT.*
 - 6: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1985.DAT.*
 - 7: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1986.DAT.*
 - 8: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1987.DAT.*
 - 9: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1988.DAT.*
 - 10: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1989.DAT.*
 - 11: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1990.DAT.*
 - 12: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1991.DAT.*
 - 13: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1992.DAT.*
 - 14: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1993.DAT.*
 - 15: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1994.DAT.*
 - 16: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1995.DAT.*
 - 17: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1996.DAT.*
 - 18: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1997.DAT.*
 - 19: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1998.DAT.*
 - 20: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1999.DAT.*
 - 21: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2000.DAT.*
 - 22: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2001.DAT.*
 - 23: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2002.DAT.*
 - 24: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	94.0	11	23	ABR80547
2	47	94.0	11	23	ABR80548
3	47	94.0	11	23	ABR80551
4	47	94.0	11	23	ABR80556
5	47	94.0	11	23	ABR80557
6	46	92.0	11	23	ABR80524
7	46	92.0	11	23	ABR80528
8	46	92.0	11	23	ABR80529
9	46	92.0	11	23	ABR80533

10	46	92.0	11	23	ABR80534	Hepatitis C virus
11	46	92.0	11	23	ABR80538	Hepatitis C virus
12	46	92.0	11	23	ABR80542	Hepatitis C virus
13	46	92.0	11	23	ABR80543	Hepatitis C virus
14	46	92.0	11	23	ABR80561	Hepatitis C virus
15	46	92.0	11	23	ABR80562	Hepatitis C virus
16	41	82.0	11	23	ABR80544	Hepatitis C virus
17	41	82.0	11	23	ABR80545	Hepatitis C virus
18	41	82.0	11	23	ABR80549	Hepatitis C virus
19	41	82.0	11	23	ABR80552	Hepatitis C virus
20	41	82.0	11	23	ABR80553	Hepatitis C virus
21	40	80.0	11	23	ABR80521	Hepatitis C virus
22	40	80.0	11	23	ABR80522	Hepatitis C virus
23	40	80.0	11	23	ABR80525	Hepatitis C virus
24	40	80.0	11	23	ABR80526	Hepatitis C virus
25	40	80.0	11	23	ABR80530	Hepatitis C virus
26	40	80.0	11	23	ABR80535	Hepatitis C virus
27	40	80.0	11	23	ABR80536	Hepatitis C virus
28	40	80.0	11	23	ABR80539	Hepatitis C virus
29	40	80.0	11	23	ABR80540	Hepatitis C virus
30	40	80.0	11	23	ABR80546	Hepatitis C virus
31	40	80.0	11	23	ABR80550	Hepatitis C virus
32	40	80.0	11	23	ABR80554	Hepatitis C virus
33	40	80.0	11	23	ABR80555	Hepatitis C virus
34	40	80.0	11	23	ABR80559	Hepatitis C virus
35	40	80.0	11	23	ABR80563	Hepatitis C virus
36	40	80.0	11	23	ABR80564	Hepatitis C virus
37	40	80.0	11	23	ABR80565	Hepatitis C virus
38	40	80.0	11	23	ABR80566	Hepatitis C virus
39	40	80.0	11	23	ABR80567	Hepatitis C virus
40	40	80.0	11	23	ABR80568	Hepatitis C virus
41	39	78.0	11	23	ABR80523	Hepatitis C virus
42	39	78.0	11	23	ABR80527	Hepatitis C virus
43	39	78.0	11	23	ABR80531	Hepatitis C virus
44	39	78.0	11	23	ABR80532	Hepatitis C virus
45	39	78.0	11	23	ABR80537	Hepatitis C virus

ALIGNMENTS

RESULT 1		
ID	ABR80547	standard; peptide; 11 AA.
XX		
AC	ABR80547;	
XX		
DT	08-OCT-2002	(first entry)
XX		
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #27.	
XX		
KW	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.	
XX		
OS	Synthetic.	
XX		
FT	Key	Location/Qualifiers
FT	Modified-site	1 /note= "N-terminal acetyl"
FT	Modified-site	6 /note= "N-terminal acetyl"
FT	Modified-site	11 /note= "C-terminal amide"
FT	Modified-site	11 /note= "C-terminal amide"
XX		
PN	W0200208251-A2.	
XX		
PD	31-JAN-2002.	
XX		
PP	19-JUL-2001; 2001WO-US23169.	
XX		
PR	21-JUL-2000; 2000US-220101P.	
XX		

PA (CORV-) CORVAS INT INC.
XX Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
XX virus protease
PS Claim 17; Page 65; 69pp; English.
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 94.0%; Score 47; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0096; Mismatches 1; Indels 0; Gaps 0;
Matches 10; Conservative 0;
QY 1 EEVVPKXGXDYS 11
1 EEVVPKXGTDYS 11
Db
RESULT 2
ABB80548
ID ABB80548 standard; peptide; 11 AA.
XX ABB80548;
AC
XX 08-OCT-2002 (first entry)
DT
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #28.
DE
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM virucide.
XX
XX Synthetic.
OS
XX Key Location/Qualifiers
FH Modified-site 1
FT /note= "N-terminal acetyl"
FT Modified-site 6
FT /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9
FT /note= "D-form residue"
FT Modified-site 11
FT /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C

PT virus protease -
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 94.0%; Score 47; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0096; Mismatches 1; Indels 0; Gaps 0;
Matches 10; Conservative 0;
QY 1 EEVVPKXGXDYS 11
1 EEVVPKXGTDYS 11
Db
RESULT 3
ABB80551
ID ABB80551 standard; peptide; 11 AA.
XX ABB80551;
AC
XX 08-OCT-2002 (first entry)
DT
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #31.
DE
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM virucide.
XX
XX Synthetic.
OS
XX Key Location/Qualifiers
FH Modified-site 1
FT /note= "N-terminal acetyl"
FT Modified-site 6
FT /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9
FT /note= "D-form residue"
FT Modified-site 11
FT /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
XX virus protease
PS Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of

RESULT 6	
ABB80524	
ID	ABB80524 standard; peptide; 11 AA.
AC	
XX	ABB80524;
DT	08-OCT-2002 (first entry)
XX	
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.
XX	
KM	Hepatitis C virus, HCV, serine protease, inhibitor; alpha-ketoamide;
KW	virucide.
XX	
OS	Synthetic.
XX	
PH	Key
FT	Modified-site
FT	/note= "N-terminal acetyl"
FT	Modified-site
FT	/note= "Normalyl carbonyl forming keto-amide linkage with
FT	residue 7"
FT	Misc-difference
FT	/note= "D-form residue"
FT	Modified-site
FT	/note= "C-terminal amide"
XX	
PN	WO200208251-A2.
XX	
PD	31-JAN-2002.
XX	
PF	19-JUL-2001; 2001WO-US23169.
XX	
PR	21-JUL-2000; 2000US-220101P.
XX	
PA	(CORV-) CORVAS INT INC.
XX	
PI	Lim-willy M, Levy OE, Brunck TK;
XX	
DR	WPI; 2002-361643/39.
XX	
FT	Novel peptide compound having hepatitis C virus protease inhibitory
PT	activity useful for treating disorders associated with hepatitis C
PT	virus protease -
XX	
PS	Claim 17; Page 64; 69pp; English.
XX	
CC	The sequence represents a peptide compound of the invention having
CC	hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC	invention are alpha-ketoamide peptide analogues. The peptides have
CC	virucide activity, and are useful for treating and in the manufacture of
CC	a medicament to treat disorders associated with HCV protease. A
CC	pharmaceutical composition comprising the peptide as an active ingredient
CC	is useful for treating disorders associated with hepatitis C virus.
XX	
SQ	Sequence 11 AA;
XX	
QY	1 EEVVPXGKDYS 11
DB	1 EEVVPXGMDYS 11
XX	
RESULT 7	
ABB80528	
ID	ABB80528 standard; peptide; 11 AA.
XX	
AC	ABB80528;
XX	
DT	08-OCT-2002 (first entry)

D8	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.
XX	
KM	Hepatitis C virus; HCV, serine protease; inhibitor; alpha-ketoamide;
KW	virocidic.
OS	Synthetic.
XX	
FH	Key Location/Qualifiers
FT	Modified-site 1 /note= "N-terminal acetyl"
FT	Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT	Misc-difference 8 /note= "D-form residue"
FT	Modified-site 11 /note= "C-terminal amide"
XX	
PX	MO200208251-A2.
XK	
PD	31-JAN-2002.
XX	
PF	19-JUL-2001; 2001WO-US23169.
PR	21-JUL-2000; 2000US-220101P.
PA	(CORV-) CORVAS INT INC.
PI	Llm-wllby M, Levy OE, Brunck TK,
DR	WPI; 2002-361643/39.
PT	Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease -
PS	Claim 17; Page 64; 69pp; English.
CC	The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have vitrucidic activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus.
CY	* 1 BEVV/PXGMDYS 11 DB 1 BEVV/PXGMDYS 11
ID	AAB80529 standard; peptide; 11 AA. AAB80529; 08-OCT-2002 (first entry) Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9. Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virocidic. Synthetic.

FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 PN MO200208251-A2.
 PD 31-JAN-2002.
 PF 19-JUL-2001; 2001MO-US23169.
 PR 21-JUL-2000; 2000US-220101P.
 PA (CORV-) CORVAS INT INC.
 PI Lim-wilby M, Levy OE, Brunck TK;
 DR WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease
 PS Claim 17; Page 64; 69pp; English.
 CC The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have C-terminal activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus.
 SQ Sequence 11 AA;
 QY 1 EEVVPKXGDYS 11
 DB 1 EEVVPKXGDYS 11
 RESULT 9
 ID ABB80533 standard; peptide; 11 AA.
 AC ABB80533;
 DT 08-OCT-2002 (first entry)
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #13.
 KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"

FT Modified-site 11 /note= "C-terminal amide"
 FT MO200208251-A2.
 PD 31-JAN-2002.
 PF 19-JUL-2001; 2001MO-US23169.
 PR 21-JUL-2000; 2000US-220101P.
 PA (CORV-) CORVAS INT INC.
 PI Lim-wilby M, Levy OE, Brunck TK;
 DR WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease
 PS Claim 17; Page 64; 69pp; English.
 CC The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have C-terminal activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus.
 SQ Sequence 11 AA;
 QY 1 EEVVPKXGDYS 11
 DB 1 EEVVPKXGDYS 11
 RESULT 10
 ID ABB80534 standard; peptide; 11 AA.
 AC ABB80534;
 DT 08-OCT-2002 (first entry)
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #14.
 KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 PN MO200208251-A2.
 PD 31-JAN-2002.
 PF 19-JUL-2001; 2001MO-US23169.

XX 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX MPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus.
XX
XX Sequence 11 AA;
XX
XX Query Match 92.0%; Score 46; DB 23; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.015;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 EEVVPXGXDYS 11
XX 1 EEVVPXGXDYS 11
XX
XX
XX RESULT 11
XX ABB80538
XX ID ABB80538 standard; peptide; 11 AA.
XX
XX ABB80538;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #18.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 1 /note= "N-terminal acetyl"
XX Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
XX residue 7"
XX
XX Misc-difference 9 /note= "D-form residue"
XX Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX MPI; 2002-361643/39.
XX

XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus.
XX
XX Sequence 11 AA;
XX
XX Query Match 92.0%; Score 46; DB 23; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.015;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 EEVVPXGXDYS 11
XX 1 EEVVPXGXDYS 11
XX
XX
XX RESULT 12
XX ABB80542
XX ID ABB80542 standard; peptide; 11 AA.
XX
XX ABB80542;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #22.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 1 /note= "N-terminal acetyl"
XX Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
XX residue 7"
XX
XX Misc-difference 8 /note= "D-form residue"
XX Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX MPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease -
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having

CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 92.0%; Score 46; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.015; Mismatches 1; Indels 0; Gaps 0;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EEVVPXGXDYS 11
1 EEVVPXGQDYS 11

RESULT 13

ABB80543 ID ABB80543 standard; peptide; 11 AA.

XX ABB80543;

DT 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #23.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
virucide.

XX Synthetic.

FT Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "N-terminyl carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

FT WO200208251-A2.

XX 31-JAN-2002.

PD 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

PI Llm-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

PT Novel peptide compound having hepatitis C virus protease inhibitory
activity useful for treating disorders associated with hepatitis C

XX Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 92.0%; Score 46; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.015; Mismatches 1; Indels 0; Gaps 0;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EEVVPXGXDYS 11
1 EEVVPXGQDYS 11

RESULT 14

ABB80561 ID ABB80561 standard; peptide; 11 AA.

XX ABB80561;

DT 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #41.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
virucide.

XX Synthetic.

FT Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "N-terminyl carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Modified-site 8 /note= "Oxymethionine"

FT Modified-site 11 /note= "C-terminal amide"

FT WO200208251-A2.

XX 31-JAN-2002.

PD 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

PI Llm-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

PT Novel peptide compound having hepatitis C virus protease inhibitory
activity useful for treating disorders associated with hepatitis C

XX Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 92.0%; Score 46; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.015; Mismatches 1; Indels 0; Gaps 0;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EEVVPXGXDYS 11
 |||||
 Db 1 EEVVPXGMDYS 11

Search completed: December 22, 2003, 17:41:02
 Job time : 32.4667 secs

RESULT 15

ABB80562
 ID ABB80562 standard; peptide; 11 AA.

XX AC ABB80562;

XX DT 08-OCT-2002 (first entry)

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #42.

XX KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Modified-site 8 /note= "Oxymethionine"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

XX W0200208251-A2.

XX PD 31-JAN-2002.

XX PF 19-JUL-2001; 2001WO-US23169.

XX PR 21-JUL-2000; 2000US-220101P.

XX PA (CORV-) CORVAS INT INC.

XX PI Lim-wilby M, Levy OE, Brunck TK;

XX DR WPI; 2002-361643/39.

XX PT Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 FT virus protease -

XX PS Claim 17; Page 65; 69pp; English.

XX CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 is useful for treating disorders associated with hepatitis C virus.

XX SQ Sequence 11 AA;

Query Match 92.0%; Score 46; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.015; 1; Indels 0; Gaps 0;

OY 1 EEVVPXGXDYS 11
 |||||
 Db 1 EEVVPXGMDYS 11

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 16:37:03 ; Search time 10.1333 Seconds
(without alignments)
45.930 Million cell updates/sec

Title: US-09-909-164-46

Perfect score: 50

Sequence: 1 BEVVPKXGDYS 11

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTCDS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	33	66.0	140	3	US-08-569-147-76 Sequence 76, Appl
2	33	66.0	140	3	US-08-569-147-82 Sequence 82, Appl
3	32	64.0	139	4	US-09-134-001C-5124 Sequence 5124, Ap
4	32	64.0	301	4	US-09-830-217-14 Sequence 14, Appl
5	31	62.0	622	2	US-08-459-146-2 Sequence 2, Appl
6	31	62.0	622	2	US-08-459-065-2 Sequence 2, Appl
7	31	62.0	739	3	US-09-413-814-86 Sequence 86, Appl
8	30.5	61.0	669	4	US-09-617-594A-2 Sequence 2, Appl
9	30	60.0	59	4	US-08-963-851-14 Sequence 14, Appl
10	30	60.0	121	4	US-09-152-060-68 Sequence 68, Appl
11	30	60.0	121	4	US-09-152-060-85 Sequence 85, Appl
12	30	60.0	122	2	US-08-879-995A-1 Sequence 1, Appl
13	30	60.0	122	3	US-09-215-096-1 Sequence 359, Appl
14	30	60.0	135	4	US-09-996-243-359 Sequence 21, Appl
15	30	60.0	231	3	US-08-926-842B-20 Sequence 20, Appl
16	30	60.0	240	3	US-08-926-842B-21 Sequence 21, Appl
17	30	60.0	241	3	US-08-834-776A-2 Sequence 2, Appl
18	30	60.0	341	3	US-08-853-948B-4 Sequence 4, Appl
19	30	60.0	341	3	US-08-853-948B-5 Sequence 5, Appl
20	30	60.0	368	4	US-09-697-367-24 Sequence 24, Appl
21	30	60.0	421	3	US-09-093-448-4 Sequence 4, Appl
22	30	60.0	421	4	US-09-813-555-4 Sequence 4, Appl
23	30	60.0	425	4	US-09-252-991A-25088 Sequence 25088, A
24	30	60.0	478	2	US-09-040-799-3 Sequence 3, Appl
25	30	60.0	478	3	US-09-093-448-1 Sequence 1, Appl
26	30	60.0	478	3	US-09-093-448-2 Sequence 2, Appl
27	30	60.0	478	3	US-09-093-448-3 Sequence 3, Appl

28	30	60.0	478	4	US-09-813-555-1 Sequence 1, Appl
29	30	60.0	478	4	US-09-813-555-2 Sequence 2, Appl
30	30	60.0	478	4	US-09-813-555-3 Sequence 3, Appl
31	30	60.0	480	4	US-09-252-991A-22164 Sequence 22164, A
32	30	60.0	1049	4	US-09-394-272-10 Sequence 10, Appl
33	30	60.0	1068	2	US-08-429-054A-11 Sequence 11, Appl
34	30	60.0	1068	3	US-09-051-341-7 Sequence 7, Appl
35	30	60.0	1068	3	US-09-394-272-8 Sequence 8, Appl
36	30	60.0	1068	4	US-09-394-272-4 Sequence 4, Appl
37	30	60.0	1081	4	US-09-394-272-11 Sequence 11, Appl
38	30	60.0	1083	4	US-09-394-272-9 Sequence 9, Appl
39	30	60.0	100	2	US-09-047-125-27 Sequence 27, Appl
40	29	58.0	100	3	US-07-736-335B-27 Sequence 27, Appl
41	29	58.0	364	4	US-09-653-375B-2 Sequence 2, Appl
42	29	58.0	381	4	US-09-134-001C-3003 Sequence 3003, Ap
43	29	58.0	493	4	US-09-411-628-10 Sequence 10, Appl
44	29	58.0	543	4	US-09-739-455-4 Sequence 4, Appl
45	29	58.0			

ALIGNMENTS

RESULT 1
US-08-569-147-76
; Sequence 76, Application US/08569147
; Patent No. 6180377
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HUMANISED ANTIBODIES
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSER: No. 6180377715, LLP
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,147
; FILING DATE: 25-March-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Trujillo, Doreen Yanko
; REGISTRATION NUMBER: 35,719
; REFERENCE/DOCKET NUMBER: CARP-0047
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULAR TYPE: protein
; US-08-569-147-76

Query Match 66.0%; Score 33; DB 3; Length 140;
Best Local Similarity 75.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPKXGDY 10
DB 122 VVPKXGDY 129

RESULT 2

US-08-569-147-82
; Sequence 82, Application US/08569147
; Patent No. 6180377
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HUMANISED ANTIBODIES
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS: 95
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 618037718, LLP
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (ERO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,147
; FILING DATE: 25-March-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Trujillo, Doreen Yalco
; REGISTRATION NUMBER: 35,719
; REFERENCE/DOCKET NUMBER: CARP-0047
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3439
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-569-147-82

Query Match 66.0%; Score 33; DB 3; Length 140;
Best Local Similarity 75.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXDY 10
Db 122 VVPXGFDY 129

RESULT 3
US-09-134-001C-5124
; Sequence 5124, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCES: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5124
; LENGTH: 139
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-5124

Query Match 64.0%; Score 32; DB 4; Length 139;
Best Local Similarity 55.6%; Pred. No. 32;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXDS 11
Db 32 VVPXGHDYN 40

RESULT 4
US-09-830-217-14
; Sequence 14, Application US/09830217
; Patent No. 6521441
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Staphylococcus aureus Genes and Polypeptides
; FILE REFERENCE: PB46PCT
; CURRENT APPLICATION NUMBER: US/09/830,217
; CURRENT FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: PCT/US99/06199
; PRIOR FILING DATE: 1999-03-18
; PRIOR APPLICATION NUMBER: 60/080,296
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/084,674
; PRIOR FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 14
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-830-217-14

Query Match 64.0%; Score 32; DB 4; Length 301;
Best Local Similarity 55.6%; Pred. No. 77;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXDS 11
Db 195 VVPXGHDYN 203

RESULT 5
US-08-459-146-2
; Sequence 2, Application US/08459146
; Patent No. 5866405
; GENERAL INFORMATION:
; APPLICANT: Choi, Gil Ho
; TITLE OF INVENTION: Genetically Engineered Transmissible
; TITLE OF INVENTION: Hypovirulence
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,146
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; APPLICATION NUMBER: US 07/832,117
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Roseman, Catherine R
; REGISTRATION NUMBER: 34,240
; REFERENCE/DOCKET NUMBER: 8589
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (201) 235-6208
TELEFAX: (201) 235-3500
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 622 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Endochia parasitica (Cryphonectria
ORGANISM: parasitica)
STRAIN: EP713
US-08-459-146-2

Query Match 62.0%; Score 31; DB 2; Length 622;
Best Local Similarity 85.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 BEVVPXG 7
Db 31 BEVVPAG 37

RESULT 6
US-08-459-065-2
Sequence 2, Application US/08459065
Patent No. 5882642
GENERAL INFORMATION:
APPLICANT: Choi, G.I. Ho
APPLICANT: Nuss, Donald Lee
TITLE OF INVENTION: Genetically Engineered Transmissible
TITLE OF INVENTION: Hypovirulence
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSER: George M. Gould, Esq., Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,065
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/832,117
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Roseman, Catherine R
REGISTRATION NUMBER: 34,240
REFERENCE/DOCKET NUMBER: 8589
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-6208
TELEFAX: (201) 235-3500
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 622 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Endochia parasitica (Cryphonectria
ORGANISM: parasitica)
STRAIN: EP713
US-08-459-065-2

Query Match 62.0%; Score 31; DB 2; Length 622;
Best Local Similarity 85.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 BEVVPXG 7
Db 31 BEVVPAG 37

RESULT 7
US-09-413-814-86
Sequence 86, Application US/09413814
Patent No. 6225064
GENERAL INFORMATION:
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
APPLICANT: Bristol-Myers Squibb, Co.
APPLICANT: Beyer, Stefan
APPLICANT: Bioecker, Helmut
APPLICANT: Brandt, Petra
APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hoffe, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
TITLE OF INVENTION: heteropolyketide compounds
FILE REFERENCE: PCT/US 99/23535
CURRENT APPLICATION NUMBER: US/09/413,814
CURRENT FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
EARLIER FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 107
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 86
LENGTH: 739
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-86

Query Match 62.0%; Score 31; DB 3; Length 739;
Best Local Similarity 62.5%; Pred. No. 3.4e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 VPXGDYS 11
Db 663 IPLGDYS 670

RESULT 8
US-09-617-594A-2
Sequence 2, Application US/09617594A
Patent No. 6541458
GENERAL INFORMATION:
APPLICANT: Audonnet, et al.
TITLE OF INVENTION: FELINE CALICIVIRUS GENES AND VACCINES IN PARTICULAR RECOMBINANT
FILE REFERENCE: 454313-3151.1
CURRENT APPLICATION NUMBER: US/09/617,594A
CURRENT FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/199,332
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: France 00 01761
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: France 99 09421
PRIOR FILING DATE: 1999-07-16
NUMBER OF SEQ ID NOS: 26
SOFTWARE: Patentin version 3.0
SEQ ID NO 2
LENGTH: 669
TYPE: PRT
ORGANISM: Feline calicivirus
US-09-617-594A-2

Query Match 61.0%; Score 30.5; DB 4; Length 669;
Best Local Similarity 63.6%; Pred. No. 3.8e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Oy 1 BEVVPXGDXS 11
Db 428 EELTPAG-DYS 437

RESULT 9
US-08-963-851-14

; Sequence 14, Application US/08963851
; Patent No. 6300116
; GENERAL INFORMATION:
; APPLICANT: VAN DER OSTEN, CLAUS
; APPLICANT: HALKIER, TORDEN
; APPLICANT: ANDERSEN, CARSTEN
; APPLICANT: BAUDITZ, PETER
; APPLICANT: HANSEN, PETER KAMP
; TITLE OF INVENTION: PROTEASE VARIANTS AND COMPOSITIONS
; FILE REFERENCE: 4946,200-US
; CURRENT APPLICATION NUMBER: US/08/963,851
; CURRENT FILING DATE: 1997-11-04
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-08-963-851-14

Query Match 60.0%; Score 30; DB 4; Length 59;
Best Local Similarity 45.5%; Pred. No. 31;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 1 BEVVPXGDXS 11
Db 38 EKHIPGLHYS 48

RESULT 10
US-09-152-060-68

; Sequence 68, Application US/09152060
; Patent No. 6448230
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: P2003P1.US
; CURRENT APPLICATION NUMBER: US/09/152,060
; CURRENT FILING DATE: 1998-09-11
; EARLIER APPLICATION NUMBER: PCT/US98/04858
; EARLIER FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: 60/040,762
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/040,710
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/050,934
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,100
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,357
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,189
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/057,765
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/048,970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/068,368
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 68

; LENGTH: 121
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-152-060-68

Query Match 60.0%; Score 30; DB 4; Length 121;
Best Local Similarity 85.7%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 BEVVPXG 7
Db 28 BEVVPXG 34

RESULT 11
US-09-152-060-85

; Sequence 85, Application US/09152060
; Patent No. 6448230
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: P2003P1.US
; CURRENT APPLICATION NUMBER: US/09/152,060
; CURRENT FILING DATE: 1998-09-11
; EARLIER APPLICATION NUMBER: PCT/US98/04858
; EARLIER FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: 60/040,762
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/040,710
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/050,934
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,100
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,357
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,189
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/057,765
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/048,970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/068,368
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 85
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (67)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (89)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-152-060-85

Query Match 60.0%; Score 30; DB 4; Length 121;
Best Local Similarity 85.7%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 BEVVPXG 7
Db 28 BEVVPXG 34

RESULT 12
US-08-879-995A-1

; Sequence 1, Application US/0887995A
; Patent No. 5985606

GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Kaser, Matthew R.
TITLE OF INVENTION: HUMAN PREPROTACHKININ B
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/879,995A
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0326 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 122 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRAITUT03
CLONE: 2109906
US-08-879-995A-1

Query Match 60.0%; Score 30; DB 2; Length 122;
Best Local Similarity 85.7%; Pred. No. 70;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EEVVPXG 7
DB 28 EEVVPXG 34

RESULT 13
US-09-215-096-1
Sequence 1, Application US/09215096
Patent No. 6008194
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Kaser, Matthew R.
TITLE OF INVENTION: HUMAN PREPROTACHKININ B
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/215,096
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/879,995
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0326 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 122 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRAITUT03
CLONE: 2109906
US-09-215-096-1

Query Match 60.0%; Score 30; DB 3; Length 122;
Best Local Similarity 85.7%; Pred. No. 70;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EEVVPXG 7
DB 28 EEVVPXG 34

RESULT 14
US-09-996-243-359
Sequence 359, Application US/09996243
Patent No. 6478825
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Batton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerltsen, Mary B.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kijavich, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas P.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
Acids Encoding the Same
FILE REFERENCE: P2730P1C13
CURRENT APPLICATION NUMBER: US/09/996,243
CURRENT FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16

PRIOR FILING DATE:	1998-06-16
PRIOR APPLICATION NUMBER:	60/089512
PRIOR FILING DATE:	1998-06-16
PRIOR APPLICATION NUMBER:	60/089514
PRIOR FILING DATE:	1998-06-16
PRIOR APPLICATION NUMBER:	60/089522
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089538
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089588
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089599
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089600
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089653
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089801
PRIOR FILING DATE:	1998-06-18
PRIOR APPLICATION NUMBER:	60/089907
PRIOR FILING DATE:	1998-06-18
PRIOR APPLICATION NUMBER:	60/089947
PRIOR FILING DATE:	1998-06-18
PRIOR APPLICATION NUMBER:	60/089948
PRIOR FILING DATE:	1998-06-19
PRIOR APPLICATION NUMBER:	60/089955
PRIOR FILING DATE:	1998-06-19
PRIOR APPLICATION NUMBER:	60/090246
PRIOR FILING DATE:	1998-06-22
PRIOR APPLICATION NUMBER:	60/090252
PRIOR FILING DATE:	1998-06-22
PRIOR APPLICATION NUMBER:	60/090254
PRIOR FILING DATE:	1998-06-22
PRIOR APPLICATION NUMBER:	60/090349
PRIOR FILING DATE:	1998-06-23
PRIOR APPLICATION NUMBER:	60/090355
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090423
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090433
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090433
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090444
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090445
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090472
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090533
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090544
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090555
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090676
PRIOR FILING DATE:	1998-06-25
PRIOR APPLICATION NUMBER:	60/090676
PRIOR FILING DATE:	1998-06-25
PRIOR APPLICATION NUMBER:	60/090696
PRIOR FILING DATE:	1998-06-25
PRIOR APPLICATION NUMBER:	60/090862
PRIOR FILING DATE:	1998-06-26

```

; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

```

```

Query Match      60.0%; Score 30; DB 4; Length 135;
Best Local Similarity 85.7%; Pred. No. 78;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy      1 EVVVPXG 7
        |||||
Db      28 EVVVPXG 34

```

```

RESULT 15
US-08-926-842B-20
; Sequence 20, Application US/08926842B
; Patent No. 6030807
; GENERAL INFORMATION:
; APPLICANT: Sa-No. 6030807ueira, Isabel
; APPLICANT: de Lancastre, Herminia
; TITLE OF INVENTION: HIGHLY REGULABLE PROMOTER FOR HETEROLOGOUS GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/926,842B
; FILING DATE: 10-SEP-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-089 N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 231 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULAR TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:

```

```

; ORGANISM: Escherichia coli
US-08-926-842B-20

```

```

Query Match      60.0%; Score 30; DB 3; Length 231;
Best Local Similarity 55.6%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

Oy      3 VVPXGXDXS 11
        :|||
Db      41 IKPSGVYDS 49

```

```

Search completed: December 22, 2003, 16:43:46
Job time : 10.1333 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 16:40:14 ; Search time 20.8667 Seconds
(without alignments)
98.451 Million cell updates/sec

Title: US-09-909-164-46
Perfect score: 50
Sequence: 1 EKVYPXGXDYS 11

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgnt2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
2: /cgnt2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgnt2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgnt2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgnt2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgnt2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
7: /cgnt2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
8: /cgnt2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgnt2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
10: /cgnt2_6/ptodata/2/pubppaa/US09B_PUBCOMB.pep:*
11: /cgnt2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
12: /cgnt2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
13: /cgnt2_6/ptodata/2/pubppaa/US10A_PUBCOMB.pep:*
14: /cgnt2_6/ptodata/2/pubppaa/US10B_PUBCOMB.pep:*
15: /cgnt2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep:*
16: /cgnt2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
17: /cgnt2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
18: /cgnt2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	68.0	283	10 US-09-738-626-4881	Sequence 4881, Ap
2	34	68.0	322	15 US-10-156-761-12418	Sequence 12418, A
3	33	66.0	153	12 US-10-029-186-32486	Sequence 32486, A
4	33	66.0	222	12 US-10-091-007-58	Sequence 58, Appl
5	33	66.0	440	11 US-09-813-408-27	Sequence 27, Appl
6	32	64.0	14	12 US-10-231-417-403	Sequence 403, App
7	32	64.0	14	12 US-10-231-417-406	Sequence 406, App
8	32	64.0	301	12 US-10-278-946-14	Sequence 14, Appl
9	32	64.0	873	12 US-10-132-350-2	Sequence 2, Appl1
10	32	64.0	873	12 US-10-132-350-4	Sequence 4, Appl1
11	32	64.0	1163	12 US-10-122-067-4	Sequence 4, Appl1
12	32	64.0	1499	12 US-10-096-534-67	Sequence 67, Appl1
13	32	64.0	1499	12 US-10-122-067-2	Sequence 2, Appl1
14	32	64.0	3472	12 US-10-029-120-4	Sequence 4, Appl1
15	32	64.0	3472	14 US-10-027-806-4	Sequence 4, Appl1

16	32	64.0	3472	14 US-10-034-623-4	Sequence 4, Appl1
17	32	64.0	3472	15 US-10-027-801-4	Sequence 4, Appl1
18	31	62.0	299	9 US-09-815-242-10697	Sequence 10697, A
19	31	62.0	653	11 US-09-820-843R-26	Sequence 26, Appl1
20	31	62.0	846	9 US-09-815-242-13904	Sequence 13904, A
21	30.5	61.0	669	15 US-10-209-507-2	Sequence 2, Appl1
22	30	60.0	7	11 US-09-909-062-1	Sequence 1, Appl1
23	30	60.0	7	11 US-09-909-062-9	Sequence 9, Appl1
24	30	60.0	7	11 US-09-909-062-130	Sequence 130, Appl
25	30	60.0	11	10 US-09-943-123-22	Sequence 22, Appl
26	30	60.0	55	10 US-09-48-080-14	Sequence 14, Appl
27	30	60.0	79	11 US-09-764-891-4601	Sequence 4601, Ap
28	30	60.0	121	9 US-09-853-161-85	Sequence 68, Appl
29	30	60.0	121	9 US-09-853-161-85	Sequence 85, Appl
30	30	60.0	121	9 US-09-852-659A-68	Sequence 68, Appl
31	30	60.0	121	9 US-09-852-659A-85	Sequence 85, Appl
32	30	60.0	121	10 US-09-852-797-68	Sequence 68, Appl
33	30	60.0	121	10 US-09-852-797-85	Sequence 85, Appl
34	30	60.0	126	12 US-10-393-840-946	Sequence 946, App
35	30	60.0	135	9 US-09-989-723-359	Sequence 359, App
36	30	60.0	135	9 US-09-989-723-359	Sequence 359, App
37	30	60.0	135	9 US-09-989-727-359	Sequence 359, App
38	30	60.0	135	9 US-09-989-727-359	Sequence 359, App
39	30	60.0	135	10 US-09-989-731-359	Sequence 359, App
40	30	60.0	135	10 US-09-989-732-359	Sequence 359, App
41	30	60.0	135	10 US-09-991-073-359	Sequence 359, App
42	30	60.0	135	10 US-09-890-442-359	Sequence 359, App
43	30	60.0	135	10 US-09-921-163-359	Sequence 359, App
44	30	60.0	135	10 US-09-993-604-359	Sequence 359, App
45	30	60.0	135	10 US-09-990-456-359	Sequence 359, App

ALIGNMENTS

RESULT 1
US-09-738-626-4881
Sequence 4881, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENO, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OR INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 4881
LENGTH: 283
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-4881
Query Match 68.0%; Score 34; DB 10; Length 283;
Best Local Similarity 75.0%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 56 VPAGADYS 63

RESULT 2

US-10-156-761-12418
; Sequence 12418, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12418
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12418

Query Match 66.0%; Score 34; DB 15; Length 322;

Best Local Similarity 66.7%; Pred. No. 61;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 BEVVPXGXD 9

Db 65 EDVVPAGAD 73

RESULT 3

US-10-029-386-32486
; Sequence 32486, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; FILE REFERENCE: AECOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32486
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF002994.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: SWISSPROT HIT: P43363, EVALU 5.00e-46
US-10-029-386-32486

Query Match 66.0%; Score 33; DB 12; Length 153;
Best Local Similarity 60.0%; Pred. No. 43;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 BEVVPXGXDY 10
Db 15 KEVVPFGHSY 24

RESULT 4

US-10-091-007-58
; Sequence 58, Application US/10091007
; Publication No. US20030170782A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics limited
; APPLICANT: Le Page, Richard W F Wells, Jeremy W
; APPLICANT: Hamillty, Sean B
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21578WO
; CURRENT APPLICATION NUMBER: US/10/091,007
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: GB 9921125.2
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 58
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-10-091-007-58

Query Match 66.0%; Score 33; DB 12; Length 222;

Best Local Similarity 50.0%; Pred. No. 65;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 1 BEVVPXGXDY 10

Db 201 KKIIVPIGFDY 210

RESULT 5

US-09-813-408-27
; Sequence 27, Application US/09813408
; Publication No. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Maris, Barry
; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides And Combinatorial
; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813,408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 27
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Aeropyrum pernix
US-09-813-408-27

Query Match 66.0%; Score 33; DB 11; Length 440;

Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 2 BEVVPXGXDY 10

Db 120 EVLPWGVGY 128

RESULT 6

US-10-231-417-403
; Sequence 403, Application US/10231417
; Publication No. US2003017681A1
; GENERAL INFORMATION:
; APPLICANT: Peng et al.
; TITLE OF INVENTION: 148 Human Secreted Proteins
; FILE REFERENCE: P2019P1

```

; CURRENT APPLICATION NUMBER: US/10/231.417
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: US/09/296,622
; PRIOR FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 619
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 403
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-231-417-403

Query Match
Best Local Similarity 64.0%; Score 32; DB 12; Length 14;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 VPKGXGYS 11
DB 1 VPCGTDYS 8

RESULT 7
US-10-231-417-406
; Sequence 406, Application US/10231417
; Publication No. US20030176681A1
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; TITLE OF INVENTION: 148 Human Secreted Proteins
; FILE REFERENCE: P2019P1
; CURRENT APPLICATION NUMBER: US/10/231.417
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: US/09/296,622
; PRIOR FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 619
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 406
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-231-417-406

Query Match
Best Local Similarity 64.0%; Score 32; DB 12; Length 14;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 VPKGXGYS 11
DB 1 VPCGTDYS 8

RESULT 8
US-10-278-946-14
; Sequence 14, Application US/10278946
; Publication No. US2003015373A1
; GENERAL INFORMATION:
; APPLICANT: Simpson et al.
; TITLE OF INVENTION: Staphylococcus aureus Genes and Polypeptides
; FILE REFERENCE: P461USD1
; CURRENT APPLICATION NUMBER: US/10/278.946
; CURRENT FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: US 09/830,217
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: PCT/US99/06199
; PRIOR FILING DATE: 1999-03-18
; PRIOR APPLICATION NUMBER: US 60/078,682
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: US 60/080,296
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: US 60/084,674
; PRIOR FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 14
```

```

; LENGTH: 301
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-278-946-14

Query Match
Best Local Similarity 64.0%; Score 32; DB 12; Length 301;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 3 VPKGXGYS 11
DB 195 IVPGRDYN 203

RESULT 9
US-10-132-350-2
; Sequence 2, Application US/10132350
; Publication No. US20030166855A1
; GENERAL INFORMATION:
; APPLICANT: Acevedo, Pedro A. Navarro
; APPLICANT: Davick, Jonathan P.
; APPLICANT: Kolomiets, Mikhailo V.
; TITLE OF INVENTION: Lipoxigenase Polynucleotides and Methods
; FILE REFERENCE: 35718/246439
; CURRENT APPLICATION NUMBER: US/10/132,350
; CURRENT FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: US 60/286,889
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/305,366
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 873
; TYPE: PRT
; ORGANISM: Zea mays
US-10-132-350-2

Query Match
Best Local Similarity 64.0%; Score 32; DB 12; Length 873;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 BEVVPKXGXDY 10
DB 340 KDVLPGVGDY 349

RESULT 10
US-10-132-350-4
; Sequence 4, Application US/10132350
; Publication No. US20030166855A1
; GENERAL INFORMATION:
; APPLICANT: Acevedo, Pedro A. Navarro
; APPLICANT: Davick, Jonathan P.
; APPLICANT: Kolomiets, Mikhailo V.
; TITLE OF INVENTION: Lipoxigenase Polynucleotides and Methods
; FILE REFERENCE: 35718/246439
; CURRENT APPLICATION NUMBER: US/10/132,350
; CURRENT FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: US 60/286,889
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/305,366
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 873
; TYPE: PRT
; ORGANISM: Zea mays
```

US-10-132-350-4

Query Match

Best Local Similarity 64.0%; Score 32; DB 12; Length 873;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;Qy 1 EEVVPXGXDYS 10
Db 340 KDVLPGVGDY 349

RESULT 11

US-10-122-067-4
; Sequence 4, Application US/10122067
; Publication No. US20030165883A1
; GENERAL INFORMATION:
; APPLICANT: RORY A.J. CURTIS
; TITLE OF INVENTION: 27091, A PHOSPHOLIPID TRANSPORTING
; FILE REFERENCE: MP101-058P1R
; CURRENT APPLICATION NUMBER: US/10/122,067
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/283434
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1163
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-122-067-4

Query Match

Best Local Similarity 63.6%; Score 32; DB 12; Length 1163;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;Qy 1 EEVVPXGXDYS 11
Db 133 EEVPRGGSVS 143

RESULT 12

US-10-096-534-67
; Sequence 67, Application US/10096534
; Publication No. US20030166887A1
; GENERAL INFORMATION:
; APPLICANT: The Brigham and Women's Hospital, Inc.
; APPLICANT: Yates, Karen
; APPLICANT: Mizuno, Shuichi
; APPLICANT: Glowacki, Julie
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF SKELETAL DEGENERATION CONDITIONS
; FILE REFERENCE: B0801/7244/KX/ERP
; CURRENT APPLICATION NUMBER: US/10/096,534
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: US 60/274,980
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 67
; LENGTH: 1499
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-096-534-67

Query Match

Best Local Similarity 64.0%; Score 32; DB 12; Length 1499;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;Qy 1 EEVVPXGXDYS 11
Db 469 EEVPRGGSVS 479

RESULT 13

US-10-122-067-2
; Sequence 2, Application US/10122067
; Publication No. US20030165883A1
; GENERAL INFORMATION:
; APPLICANT: RORY A.J. CURTIS
; TITLE OF INVENTION: 27091, A PHOSPHOLIPID TRANSPORTING
; FILE REFERENCE: MP101-058P1R
; CURRENT APPLICATION NUMBER: US/10/122,067
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/283434
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1499
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-122-067-2

Query Match

Best Local Similarity 64.0%; Score 32; DB 12; Length 1499;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;Qy 1 EEVVPXGXDYS 11
Db 469 EEVPRGGSVS 479

RESULT 14

US-10-029-120-4
; Sequence 4, Application US/10029120
; Publication No. US20030175708A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM
; FILE REFERENCE: DCCRP.002A
; CURRENT APPLICATION NUMBER: US/10/029,120
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-029-120-4

Query Match

Best Local Similarity 45.5%; Score 32; DB 12; Length 3472;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;Qy 1 EEVVPXGXDYS 11
Db 2294 EDVIPRGISFS 2304

RESULT 15

US-10-027-806-4
; Sequence 4, Application US/10027806
; Publication No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM
; FILE REFERENCE: DCCRP.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; PRIOR FILING DATE: 2001-12-21

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-806-4

Query Match 64.0%; Score 32; DB 14; Length 3472;
Best Local Similarity 45.5%; Pred. NO. 2.2e+03;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 BRVVPXGXDYS 11
DB 2294 EDVIRGISFS 2304

Search completed: December 22, 2003, 17:32:45
Job time : 20.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:24:36 ; Search time 9.06667 Seconds

(without alignments)
116.675 Million cell updates/sec

Title: US-09-909-164-46

Perfect score: 50

Sequence: 1 BEVVPKXGDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76:*

1: Pir1:*

2: Pir2:*

3: Pir3:*

4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	72.0	363	2	D69551
2	35	70.0	290	2	AG3104
3	35	70.0	290	2	D98182
4	35	70.0	587	2	P81138
5	35	70.0	906	2	T48898
6	35	70.0	908	2	T48899
7	34	68.0	102	2	A42452
8	34	68.0	156	2	S54619
9	34	68.0	247	2	A96001
10	34	68.0	257	2	A96546
11	34	68.0	394	2	P82491
12	34	68.0	433	2	H87660
13	34	68.0	2747	2	B49132
14	33	66.0	124	1	VKLJST
15	33	66.0	165	2	AG1272
16	33	66.0	165	2	AH1635
17	33	66.0	196	2	AD0454
18	33	66.0	225	2	S57810
19	33	66.0	327	2	S40753
20	33	66.0	421	1	D8RTCM
21	33	66.0	440	2	H72784
22	33	66.0	1028	2	AP3286
23	33	66.0	1088	2	D82246
24	33	66.0	1150	2	T20173
25	32	64.0	99	2	S00210
26	32	64.0	155	2	S38255
27	32	64.0	168	2	S58208
28	32	64.0	196	2	AI0931
29	32	64.0	301	2	P89957

30	32	64.0	307	2	P84330	hypothetical prote
31	32	64.0	314	2	AH1912	hypothetical prote
32	32	64.0	357	1	G69290	probable hexosyltr
33	32	64.0	366	2	G69350	L-lactate dehydrog
34	32	64.0	425	2	T24111	hypothetical prote
35	32	64.0	427	2	F64064	colB protein - Hae
36	32	64.0	565	2	B86665	ABC transporter AT
37	32	64.0	632	2	H84350	oligopeptidase (tm
38	32	64.0	672	2	G88651	protein B0212.3 [1
39	32	64.0	1474	2	F69009	probable membrane
40	32	64.0	3472	2	T13108	hypothetical 367K
41	32	64.0	6658	2	T13931	projectin - fruit
42	31	62.0	117	2	A69487	response regulator
43	31	62.0	202	2	H97247	S-adenosylmethioni
44	31	62.0	233	2	E72350	conserved hypothet
45	31	62.0	296	2	F72745	hypothetical prote

ALIGNMENTS

RESULT 1

D69551 conserved hypothetical protein AF2411 - Archaeoglobus fulgidus

C/Species: Archaeoglobus fulgidus

C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999

C/Accession: D69551

R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.R.; Ketchum, K.A.; Dodson, J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirness, E.

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A/Authors: Usterback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, Smith, H.O.; Moese, C.R.; Venter, J.C.

A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch

A/Reference number: A69250; MUID:98049343; PMID:9389475

A/Accession: D69551

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-363 <KLE>

A/Cross-references: GB:AE001109; GB:AE000782; MID:92669432; PIDN:AMB91255.1; PID:92651

Query Match 72.0%; Score 36; DB 2; Length 363;
Best Local Similarity 54.5%; Pred. NO. 9.1;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Cy 1 BEVVPKXGDYS 11

DB 120 ENIVPYGIDFS 130

RESULT 2

AG3104 6-O-methylguanine-DNA methyltransferase [imported] - Agrobacterium tumefaciens (strain

C/Species: Agrobacterium tumefaciens

C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C/Accession: AG3104

R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, Erase, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl

Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Kreepan, W.; Perry, M.; Gordon-Kam

eter, B.W.

A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A/Reference number: AB2577; MUID:21608550; PMID:11743193

A/Accession: AG3104

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-290 <KUR>

A/Cross-references: GB:AE006689; PIDN:AA145253.1; PID:917742937; GSPDB:GN00107

A/Experimental source: strain C58 (Dupont)

C/Genetics:

A/Gene: ada

A/Map position: linear chromosome

Query Match 70.0%; Score 35; DB 2; Length 290;
 Best Local Similarity 50.0%; Pred. No. 12;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 EEVVPXGXDY 10
 ||:|||||
 Db 9 EDIRPIGSDY 18

RESULT 3

D98182

06-methylguanine-DNA methyltransferase PA2118 [imported] - Agrobacterium tumefaciens (str
 C:Species: Agrobacterium tumefaciens
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002

C:Accession: D98182
 R:Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
 A:Reference number: A97359; MUID:21608551; PMID:11743194
 A:Accession: D98182
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-290 <CDR>
 A:Cross-references: GB:AE007870; PIDN:AAK8982.1; PID:G15158766; GSPDB:GN00170
 C:Genetics:
 A:Gene: AGR_L_818
 A:Map position: linear chromosome

Query Match 70.0%; Score 35; DB 2; Length 290;
 Best Local Similarity 50.0%; Pred. No. 12;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 EEVVPXGXDY 10
 ||:|||||
 Db 9 EDIRPIGSDY 18

RESULT 4

F81138

succinate dehydrogenase, flavoprotein chain NMB0950 [imported] - Neisseria meningitidis
 C:Species: Neisseria meningitidis
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C:Accession: F81138
 R:Jettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; ri, H.; Qiu, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizze, M. Science 287, 1809-1815, 2000
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; V. A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
 A:Reference number: A81000; MUID:20175755; PMID:10710307
 A:Accession: F81138
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-587 <TER>
 A:Cross-references: GB:AE002446; GB:AE002098; NID:G7226185; PIDN:AA41356.1; PID:G722618
 A:Experimental source: serogroup B, strain MC58
 C:Genetics:
 A:Gene: NMB0950
 C:Superfamily: fumarate reductase flavoprotein; 3-oxosteroid 1-dehydrogenase homology; 4

Query Match 70.0%; Score 35; DB 2; Length 587;
 Best Local Similarity 70.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 EEVVPXGXDY 10
 ||:|||||
 Db 366 EVVVPQGS DY 375

RESULT 5

T48898

disease resistance protein RPP8 [validated] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 21-Jul-2000
 C:Accession: T48898
 R:McDowell, J.M.; Dhandaaydam, M.; Long, T.A.; Aarts, M.G.; Goff, S.; Holub, E.B.; Dan Plant Cell 10, 1861-1874, 1998
 A:Title: Intragenic recombination and diversifying selection contribute to the evolution of the Arabidopsis thaliana RPP8 gene
 A:Reference number: 224999; MUID:99030193; PMID:9811794
 A:Accession: T48898
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-906 <MCD>
 A:Cross-references: EMBL:AF089710; NID:G3928861; PIDN:AA083165.1; PID:G3928862
 A:Experimental source: Landsberg erecta
 C:Genetics:
 A:Gene: RPP8
 A:Introns: 293/1; 342/1
 C:Function:
 A:Description: promotes resistance to Peronospora parasitica

Query Match 70.0%; Score 35; DB 2; Length 906;
 Best Local Similarity 60.0%; Pred. No. 40;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 EEVVPXGXDY 10
 ||:|||||
 Db 881 EKLVPQGS DY 890

RESULT 6

T48899

disease resistance protein rpp8 [similarity] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 21-Jul-2000
 C:Accession: T48899
 R:McDowell, J.M.; Dhandaaydam, M.; Long, T.A.; Aarts, M.G.; Goff, S.; Holub, E.B.; Dan Plant Cell 10, 1861-1874, 1998
 A:Title: Intragenic recombination and diversifying selection contribute to the evolution of the Arabidopsis thaliana RPP8 gene
 A:Reference number: 224999; MUID:99030193; PMID:9811794
 A:Accession: T48899
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-906 <MCD>
 A:Cross-references: EMBL:AF089711; NID:G3901293; PIDN:AA078631.1; PID:G3901294
 A:Experimental source: Columbia
 C:Genetics:
 A:Gene: rpp8
 A:Introns: 293/1; 342/1
 C:Function:
 A:Description: susceptible allele of a gene that promotes resistance to Peronospora p

Query Match 70.0%; Score 35; DB 2; Length 908;
 Best Local Similarity 60.0%; Pred. No. 40;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 EEVVPXGXDY 10
 ||:|||||
 Db 883 EKLVPQGS DY 892

RESULT 7

A42452

V1 protein - tobacco yellow dwarf virus (strain Australia)

C:Species: tobacco yellow dwarf virus
 C:Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
 C:Accession: A42452
 R:Morrice, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.B. Virology 187, 633-642, 1992
 A:Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow dwarf virus
 A:Reference number: A42452; MUID:9218538; PMID:1546458
 A:Accession: A42452
 A:Molecule type: DNA
 A:Residues: 1-102 <MOR>

A:Cross-references: GB:M01103; NID:G335283; PID:AAA47947.1; PID:G335284

Query Match 68.0%; Score 34; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 6;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVFXGXDYS 11
DB 7 QVPSGIVYS 16

RESULT 8

SS4619 hypothetical protein YOR013w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein C2612; hypothetical protein YOL03.3

C:Species: Saccharomyces cerevisiae

C:Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002

C:Accession: SS4619; S66879

R:de Haan, M.; Maarse, A.C.; Grivell, L.A.

submitted to the EMBL Data Library, May 1995

A:Reference number: S54617

A:Accession: S54619

A:Molecule type: DNA

A:Residues: 1-156 <DEH>

A:Cross-references: EMBL:X87331; NID:G1041652; PID:CAA60762.1; PID:G829123

R:de Haan, M.; Grivell, L.A.; Maarse, A.C.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S66877

A:Accession: S66879

A:Molecule type: DNA

A:Residues: 1-156 <DEH>

A:Cross-references: EMBL:Z74920; NID:G1420109; PID:CAA92201.1; PID:G1420111; MIPS:YOR01

A:Experimental source: strain S288C

C:Genetics:

A:Cross-references: SGD:S0005533

A:Map position: 15R

C:Superfamily: hypothetical protein YOR013w

Query Match 68.0%; Score 34; DB 2; Length 156;

Best Local Similarity 66.7%; Pred. No. 9.6;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVFXGXDYS 10
DB 50 EVVPLGMDY 58

RESULT 9

A96001 conserved hypothetical protein, homolog to osmotically inducible sensory protein SMC22-1

C:Species: Stenobothrium mellicol

C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: A96001

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernat

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-fixing endo

A:Reference number: A95842; MUID:21396508; PMID:11481431

A:Accession: A96001

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-247 <KUR>

A:Cross-references: GB:AL591985; PIDN:CAC49673.1; PID:G15141160; GSPDB:GN00167

R:Galibert, F.; Finan, T.M.; Long, S.R.; Punter, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

pela, D.; Chain, P.; Cowle, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kise, B.; Komp, C.; Lelaure,

hebut, P.; Vandenberg, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A:Title: The composite genome of the legume symbiont Stenobothrium mellicol.

A:Reference number: A96039; MUID:21368234; PMID:11474104

C:Contents: annotation

C:Genetics:

A:Gene: SMO21444
A:Genome: Plasmid

Query Match 68.0%; Score 34; DB 2; Length 247;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVFXGXDYS 10
DB 48 EDVSPRGADY 57

RESULT 10

A96546

unknown protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: A96546

R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Aloni

ansen, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Lin, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luross, J.S.; Maitl, R.; Marzla,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallor

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: A96546

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-257 <STO>

A:Cross-references: GB:AE005173; NID:G11094688; PIDN:AAG29624.1; GSPDB:GN00141

C:Genetics:

A:Gene: P8A12.12

A:Map position: 1

Query Match 68.0%; Score 34; DB 2; Length 257;

Best Local Similarity 54.5%; Pred. No. 17;

Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVFXGXDYS 11
DB 217 EELLKAGADYS 227

RESULT 11

F82491

ferrialdophore reductase VCA0183 [imported] - Vibrio cholerae (strain N16961 serogrc

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: F82491

R:Haidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers,

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: F82491

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1394 <HET>

A:Cross-references: GB:AE004358; GB:AE003853; NID:G9657566; PIDN:AAF96096.1; GSPDB:GN0

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VCA0183

A:Map position: 2

C:Superfamily: flavohemoglobin; cytochrome-b5; reductase homology; globin homology

Query Match 68.0%; Score 34; DB 2; Length 394;

Best Local Similarity 66.7%; Pred. No. 27;

Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 EVVFXGXDY 10
DB 194 EVVPEGSDY 202

RESULT 12

H87660
peptidoglycan-binding protein, probable [imported] - Caulobacter crescentus
C/Species: Caulobacter crescentus
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C/Accession: H87660
R/Nerman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Neilson, K.E.; Bisen, J.; Heidelberg, J.
B.; Labb, M.T.; Debay, R.T.; Dodson, R.J.; Durkin, A.S.; Gwin, M.L.; Haft, D.H.; Klot
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A/Title: Complete Genome Sequence of Caulobacter crescentus.
A/Reference number: A87249; MUID:21173698; PMID:11259647
A/Accession: H87660
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-433 <STO>
A/Cross-references: GB:AE005673; NID:G13425020; PIDN:AAK25284.1; GSPDB:GN00148
C/Genetics:
A/Gene: CC3322

Query Match 68.0%; Score 34; DB 2; Length 433;
Best Local Similarity 54.5%; Pred. No. 29;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 BEVVPXGXDYS 11
DB 266 EVIIPGPDYS 276

RESULT 13

B49132
fat facets (fat) splice form 1 - fruit fly (Drosophila melanogaster)
C/Species: Drosophila melanogaster
C/Date: 19-Dec-1993 #sequence_revision 25-Apr-1997 #text_change 01-Dec-2000
C/Accession: B49132; A49132
R/Fischer-Vize, J.A.; Rubin, G.M.; Lehmann, R.
Development 116, 985-1000, 1992
A/Title: The fat facets gene is required for Drosophila eye and embryo development.
A/Reference number: A49132; MUID:93202020; PMID:1295747
A/Accession: B49132
A/Contents: isogenic st
A/Status: preliminary
A/Molecule type: nucleic acid
A/Residues: 1-2747 <RIS>
A/Cross-references: GB:L04959; NID:G157411; PIDN:AAFO1345.1; PID:G6013474
A/Note: sequence inconsistent with the nucleotide translation
A/Note: sequence extracted from NCBI backbone (NCBIN:127836, NCBIN:129008, NCBIP:127839)
A/Accession: A49132
A/Status: preliminary
A/Molecule type: nucleic acid
A/Residues: 1-2704, 'VT', 2707, 'ANNV', <RT2>
A/Cross-references: GB:L04958; NID:G157410; PIDN:AAFO1346.1; PID:G6013475
A/Note: sequence extracted from NCBI backbone (NCBIN:127836, NCBIN:129008, NCBIP:129029)
C/Keywords: alternative splicing

Query Match 68.0%; Score 34; DB 2; Length 2747;
Best Local Similarity 54.5%; Pred. No. 2,2e+02;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 BEVVPXGXDYS 11
DB 1394 EVIYVPGDYS 1404

RESULT 14

VRLJ51
trans-regulatory splicing protein - simian immunodeficiency virus SIVcpz

N/Alternate names: anti-repression trans-activator; art protein; rev protein; trs prot
C/Species: simian immunodeficiency virus SIVcpz
A/Note: host Pan troglodytes (chimpanzee)
C/Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1993
C/Accession: S09988
R/Huet, T.; Cheynier, R.; Meyershan, A.; Roelants, G.; Wain-Hobson, S.
Nature 345, 356-359, 1990

A/Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.
A/Reference number: S09983; MUID:90259077; PMID:2188136
A/Accession: S09988
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-124 <HBE>
A/Cross-references: EMBL:X52154; NID:G58866; PIDN:CAA36405.1; PID:G763085
C/Genetics:
A/Gene: rev, trs, art
A/Intons: 27/1

C/Superfamily: AIDS trans-regulatory splicing protein
C/Keywords: AIDS; immunodeficiency; splicing protein; transcription regulation

Query Match 66.0%; Score 33; DB 1; Length 124;
Best Local Similarity 60.0%; Pred. No. 12;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 EVVFXGXDYS 11
DB 107 EVVPAAGNYS 116

RESULT 15

AG1272
thiol peroxidases homolog lmo1583 [imported] - Listeria monocytogenes (strain EGD-e)
C/Species: Listeria monocytogenes
C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 17-May-2002
C/Accession: AG1272
R/Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloesch
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurget, O.; Eutian, K.D.; Feihl,
J.; Jones, L.M.; Karet, U.
Science 294, 849-852, 2001
A/Authors: Kretz, T.; Kuhn, M.; Kunst, F.; Kurapkak, G.; Madeno, E.; Matournam, A.;
ok, C.; Schluter, T.; Simoes, N.; Tietz, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlan
A/Title: Comparative genomics of Listeria species.
A/Reference number: AB1077; MUID:21537279; PMID:11679669
A/Accession: AG1272
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-165 <GIA>
A/Cross-references: GB:NC_003210; PIDN:CA099661.1; PID:G16411012; GSPDB:GN00177
A/Experimental source: strain EGD-e
C/Genetics:
A/Gene: lmo1583
C/Superfamily: thioredoxin peroxidase

Query Match 66.0%; Score 33; DB 2; Length 165;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 EVVFXGXDY 10
DB 144 EVVPEGSDH 152

Search completed: December 22, 2003, 17:45:01
Job time : 9.06667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 16:43:51 ; Search time 4.6 Seconds

(without alignments)
112.455 Million cell updates/sec

Title: US-09-909-164-46
Perfect score: 50
Sequence: 1 EEVVPXGXDXS 11

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	35	70.0	RBL4_ARATH	Q9FJK8 arabidopsis
2	35	70.0	RPP8_ARATH	Q8W4J9 arabidopsis
3	35	70.0	RPP8_ARATH	P59584 arabidopsis
4	35	70.0	CARB_FUSON	O8FG86 fusobacteri
5	34	68.0	Y1LK_TYDVA	P31619 tobacco yel
6	34	68.0	HMPA_VIBCH	Q9KMY3 vibrio chol
7	34	68.0	RAP_BROME	P58824 drosophila
8	33	66.0	REV_STVCC	P17280 chimpanzee
9	33	66.0	TPX_LISIN	Q9Y6U8 listeria in
10	33	66.0	TPX_LISIN	Q8Y6U8 listeria in
11	33	66.0	YK14_CABEL	P08503 ratuus nov
12	33	66.0	ACDM_RAT	P08503 ratuus nov
13	33	66.0	STR_STRP3	Q8K5J2 streptococ
14	33	66.0	STR_STRP8	O8K222 streptococ
15	33	66.0	STR_STRP	Q9KX15 streptococ
16	32	64.0	PLAS_ORVSA	P20423 cryza savi
17	32	64.0	PLAS_HORYU	P08248 hordeum vul
18	32	64.0	PLAT_POPI	P11970 populus nig
19	32	64.0	YVAG_SALTY	Q91912 salmonella
20	32	64.0	YVAG_SALTY	Q91912 salmonella
21	32	64.0	ALOC_HUMAN	O60312 homo sapien
22	32	64.0	CHAC_HUMAN	O9KX17 homo sapien
23	31	62.0	HIS9_THENA	O9KX17 homo sapien
24	31	62.0	YHAI_CRYPA	P10941 cyphonectr
25	31	62.0	ACDM_HUMAN	P11310 homo sapien
26	31	62.0	ACDM_PIG	P41367 sus scrofa
27	31	62.0	TR2M_PANAY	Q47861 pantoea agy
28	30	60.0	PLAS_DAUCA	P20422 dancus caro
29	30	60.0	TKMK_HUMAN	O9KX17 homo sapien
30	30	60.0	Y149_ARCFU	O28330 archaeoglob
31	30	60.0	HES3_RAT	Q04667 ratuus nov
32	30	60.0	CTC_BACSU	P41994 bacillus su
33	30	60.0	ARAD_ECOLI	P08203 escherichia

34	30	60.0	231	1	ARAD_SALTY	P06190 salmonella
35	30	60.0	232	1	SCOA_HELPJ	Q92163 helicobacte
36	30	60.0	232	1	SCOA_HELPJ	P56006 helicobacte
37	30	60.0	259	1	OVUH_TYMST	P06308 tymsea sca
38	30	60.0	421	1	AMP2_YEAST	P38174 saccharomyc
39	30	60.0	421	1	ECB2_HALEL	O52250 halomonas e
40	30	60.0	423	1	ECB2_HALEL	O9Z6U7 halomonas e
41	30	60.0	430	1	FOLC_BACSU	O05865 bacillus su
42	30	60.0	457	1	2185_HUMAN	O15231 homo sapien
43	30	60.0	459	1	LEFT_KLUTA	P53998 kluyveromyc
44	30	60.0	478	1	AMP2_HUMAN	P50579 homo sapien
45	30	60.0	478	1	AMP2_MOUSE	O08663 mus musculu

ALIGNMENTS

RESULT 1
RBL4_ARATH STANDARD; PRT; 908 AA.
ID RBL4_ARATH
AC Q9FJK8
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable disease resistance RPP8-like protein 4.
GN RPP8L4 OR AT5G46620 OR K1SN18.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyle; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STPAIN-CV, Columbia;
RK MEDLINE=9087489; Pubmed=9872454;
RA Nakamura Y., Sato S., Asamizu E., Kaneo T., Kotani H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
RT Sequence features of the regions of 1,013,767 bp covered by sixteen
RT physically assigned P1 and TNC clones.";
RL DNA Res. 5:297-308(1998).
[2]
RC SEQUENCE FROM N.A.
RC STPAIN-CV, Columbia;
RK MEDLINE=9087489; Pubmed=9872454;
RA Nakamura Y., Sato S., Asamizu E., Kaneo T., Kotani H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana full-length cDNA.";
RT Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Potential disease resistance protein.
CC -1- DOMAIN: The LRR repeats probably act as specificity determinant of
CC pathogen recognition (by similarity).
CC -1- SIMILARITY: Belongs to the disease resistance NB-LRR family.
CC -1- SIMILARITY: RPP8/HRT subfamily.
CC -1- SIMILARITY: Contains 3 leucine-rich (LRR) repeats.
CC -1- SIMILARITY: Contains 1 NB-ARC domain.
CC -1- DATABASE: NAME=NIB-LRRS;
CC NOT3=functional and comparative genomics of disease resistance gene
CC homologs;
CC WWW=http://niblrrs.ucdavis.edu".
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
DR EMBL, AB015468; BAB10695.1; -;
DR EMBL, AK117163; BAC41841.1; -;
DR InterPro, IPR000767; Disease_resist.
DR InterPro, IPR001611; LRR.

DR InterPro: IPR002182; NB-ARC.
 DR Pfam: PF00560; LRR; 2.
 DR Pfam: PF00931; NB-ARC; 1.
 DR PRINTS: PR00364; DISEASERIST.
 KM Plant defense; ATP-binding; Repeat; Leucine-rich repeat.
 FT DOMAIN 10 45
 FT REPEAT 146 459 NB-ARC.
 FT REPEAT 575 599 LRR 1.
 FT REPEAT 600 623 LRR 2.
 FT REPEAT 842 867 LRR 3.
 FT NP_BIND 192 199 ATP (POTENTIAL).
 SQ SEQUENCE 908 AA; 104448 MW; 3111991817239693 CRC64;
 Query Match 70.0%; Score 35; DB 1; Length 908;
 Best Local Similarity 60.0%; Pred. No. 16;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Cy 1 BEVVPXGXDY 10
 Db 883 EKLVPGGEDY 892
 RESULT 2
 RPP8_ARATH STANDARD; PRT; 908 AA.
 ID RPP8_ARATH STANDARD; PRT; 908 AA.
 AC Q8W4J9; Q8W4J9; Q9ZSY3; Q9ZSY4;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Disease resistance protein RPP8 (Resistance to Peronospora parasitica
 protein 8).
 GN RPP8 OR HRT OR AT5G43470 OR MWP20.19.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucotsids II; Brassicales; Brassicaceae; Arabidopsids.
 CX NCBI_TaxId=3702;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, MUTANTS RPP8-1, RPP8-2 AND
 RPP8-3, AND VARIANTS.
 RC STRAIN=cv. Columbia, and cv. Landsberg erecta;
 RX MEDLINE=98030193; PubMed=9811794;
 RA Houch J.M., Dhandaydham M., Long T.A., Aarts M.G.M., Goff S.,
 "Intergenic recombination and diversifying selection contribute to the
 evolution of downy mildew resistance at the RPP8 locus of
 Arabidopsis";
 RL Plant Cell 10:1861-1874(1998).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS.
 RC STRAIN=cv. Di-17;
 RX MEDLINE=20271766; PubMed=10810142;
 RA Coley M.B., Patilrana S., Wu H.-J., Kachroo P., Kleesig D.F.;
 "Members of the Arabidopsis HRT/RPP8 family of resistance genes confer
 resistance to both viral and oomycete pathogens";
 RL Plant Cell 12:663-676(2000).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20181125; PubMed=10718197;
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
 Tabata S.;
 "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
 features of the regions of 3,076,755 bp covered by sixty P1 and TAC
 clones";
 RL DNA Res. 7:31-63(2000).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=cv. Columbia;
 RA Shinzaki K., Davis R.W., Becker J.R., Theologis A.;
 "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 SSP consortium (Salk/Stanford/PGSC)";
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

RN [5]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC STRAIN=cv. Columbia;
 RA Seki M., Iida K., Sato M., Sakurai T., Akiyama K., Ishida J.,
 RA Nakajima M., Enju A., Kamaya A., Narusaka M., Carninci P., Kawai J.,
 RA Hayashizaki Y., Shinozaki K.;
 "Arabidopsis thaliana full-length cDNA";
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP INTERACTION WITH TIP.
 RX MEDLINE=20496823; PubMed=11041886;
 RA Ren T., Qu P., Morris T.J.;
 "HRT gene function requires interaction between a NAC protein and
 viral capsid protein to confer resistance to turnip crinkle virus";
 RL Plant Cell 12:1917-1926(2000).
 CC -1- FUNCTION: Disease resistance protein. Resistance proteins guard
 the plant against pathogens that contain an appropriate avirulence
 protein via an indirect interaction with this avirulence protein.
 That triggers a defense system including the hypersensitive
 response, which restricts the pathogen growth. The interaction
 with TIP (TICV-interacting protein) may be essential for the
 recognition of the avirulence proteins, and the triggering of the
 defense response.
 CC -1- SUBUNIT: Interacts with the NAC protein TIP.
 CC -1- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=2;
 Name=1;
 IsoId=Q8W4J9-1; Sequence=Displayed;
 Name=2;
 IsoId=Q8W4J9-2; Sequence=VSP_007171, VSP_007172;
 Note=Has been shown to exist only in cv. Columbia so far;
 CC -1- DOMAIN: The LRR repeats probably act as specificity determinant of
 pathogen recognition.
 CC -1- POLYMORPHISM: The strong polymorphisms present in cv. Di-17 and
 cv. Columbia are probably due to an unequal crossing-over between
 the highly related RPP8 and RPP8A genes present in cv. Landsberg
 erecta. Such variations probably modify the specificity of
 pathogen recognition.
 CC -1- MISCELLANEOUS: In cv. Columbia and cv. Landsberg erecta, RPP8
 specifically recognizes the Emco5 avirulence protein from
 Peronospora parasitica, while it is not the case in cv. Di-17,
 where it confers resistance to Turnip Crinkle Virus upon
 recognition of the viral capsid protein.
 CC -1- SIMILARITY: Belongs to the disease resistance NB-LRR family.
 CC -1- RPP8/HRT subfamily.
 CC -1- SIMILARITY: Contains 2 leucine-rich (LRR) repeats.
 CC -1- SIMILARITY: Contains 1 NB-ARC domain.
 CC -1- DATABASE: NAME-NIB-LRRS;
 NOTE=Functional and comparative genomics of disease resistance gene
 homologs;
 WWW="http://niblrrs.ucdavis.edu".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF089710; AAC83165.1; -
 DR EMBL: AF089711; AAC78651.1; -
 DR EMBL: AF234114; AAF36987.1; -
 DR EMBL: AB025638; BAA97426.1; -
 DR EMBL: AY062514; AAL32592.1; -
 DR EMBL: AK118862; BAC43449.1; -
 DR InterPro: IPR000767; Disease_resistc.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR002182; NB-ARC.
 DR Pfam: PF00560; LRR; 2.
 DR Pfam: PF00931; NB-ARC; 1.
 DR PRINTS: PR00364; DISEASERIST.
 KM Plant defense; ATP-binding; Repeat; Leucine-rich repeat;

Alternative splicing; Polymorphism.
 KM DOMAIN 10 45
 FT DOMAIN 146 459
 FT REPEAT 600 623
 FT REPEAT 842 867
 FT NP BIND 192 199
 FT VARSPLIC 294 308
 FT VARSPLIC 309 908
 FT VARSPLIC 27 31
 FT VARSPLIC 29 29
 FT VARSPLIC 85 87
 FT VARSPLIC 88 91
 FT VARSPLIC 100 100
 FT VARSPLIC 121 121
 FT VARSPLIC 129 129
 FT VARSPLIC 133 133
 FT VARSPLIC 138 138
 FT VARSPLIC 173 173
 FT VARSPLIC 177 177
 FT VARSPLIC 253 253
 FT VARSPLIC 257 257
 FT VARSPLIC 265 265
 FT VARSPLIC 270 270
 FT VARSPLIC 283 283
 FT VARSPLIC 369 369
 FT VARSPLIC 387 387
 FT VARSPLIC 399 399
 FT VARSPLIC 426 426
 FT VARSPLIC 429 429
 FT VARSPLIC 436 446
 FT VARSPLIC 438 443
 FT VARSPLIC 450 450
 FT VARSPLIC 460 460
 FT VARSPLIC 480 486
 FT VARSPLIC 480 481
 FT VARSPLIC 485 486
 FT VARSPLIC 489 489
 FT VARSPLIC 514 514
 FT VARSPLIC 518 518
 FT VARSPLIC 519 519
 FT VARSPLIC 528 528
 FT VARSPLIC 537 537
 FT VARSPLIC 542 542
 FT VARSPLIC 543 543
 FT VARSPLIC 547 550
 FT VARSPLIC 549 549
 FT VARSPLIC 558 555
 FT VARSPLIC 558 560
 FT VARSPLIC 564 564
 FT VARSPLIC 584 584
 FT VARSPLIC 595 595
 FT VARSPLIC 609 612
 FT VARSPLIC 630 635
 FT VARSPLIC 632 636
 FT VARSPLIC 650 650
 FT VARSPLIC 653 654
 FT VARSPLIC 659 659
 FT VARSPLIC 660 660

Leucine zipper.
 LRR 1.
 LRR 2.
 ATP (POTENTIAL).
 WKMLTSTRNCGVGH -> ELMYTHALFILMS (in isoform 2).
 /Frida-VSP.007171.
 Missing (in isoform 2).
 /Frida-VSP.007172.
 IDGL -> VDEQI (in cv. Landsberg erecta).
 SG -> B (in cv. Di-17).
 SGK -> RGR (in cv. Di-17 and cv. Landsberg erecta).
 CKGV -> EKI (in cv. Landsberg erecta).
 C -> R (in cv. Landsberg erecta).
 B -> D (in cv. Landsberg erecta).
 F -> L (in cv. Di-17).
 Q -> QO (in cv. Di-17).
 G -> V (in cv. Landsberg erecta).
 K -> T (in cv. Di-17).
 G -> S (in cv. Di-17).
 Y -> S (in cv. Di-17).
 R -> P (in cv. Di-17).
 A -> T (in cv. Di-17).
 R -> K (in cv. Landsberg erecta).
 V -> L (in cv. Di-17).
 V -> R (in cv. Di-17).
 A -> V (in cv. Di-17 and cv. Landsberg erecta).
 F -> S (in cv. Di-17 and cv. Landsberg erecta).
 M -> C (in cv. Di-17 and cv. Landsberg erecta).
 C -> R (in cv. Di-17 and cv. Landsberg erecta).
 N -> F (in cv. Di-17 and cv. Landsberg erecta).
 DSEIYSLF -> YKISAYDLFN (in cv. Landsberg erecta).
 BISTYS -> KITQIE (in cv. Di-17).
 A -> V (in cv. Landsberg erecta).
 B -> Q (in cv. Landsberg erecta).
 DNYSMQ -> NKILRPH (in cv. Di-17).
 WQ -> NR (in cv. Landsberg erecta).
 WQ -> SH (in cv. Landsberg erecta).
 Y -> N (in cv. Landsberg erecta).
 I -> K (in cv. Di-17 and cv. Landsberg erecta).
 C -> S (in cv. Di-17 and cv. Landsberg erecta).
 T -> I (in cv. Di-17).
 S -> R (in cv. Di-17).
 S -> R (in cv. Di-17).
 H -> Q (in cv. Di-17).
 I -> L (in cv. Landsberg erecta).
 KNKT -> RNAA (in cv. Di-17).
 K -> N (in cv. Landsberg erecta).
 PRFEDY -> WDEDRG (in cv. Landsberg erecta).
 PRP -> GREK (in cv. Di-17).
 Y -> F (in cv. Di-17).
 S -> Y (in cv. Landsberg erecta).
 C -> S (in cv. Di-17 and cv. Landsberg erecta).
 YEA -> FLAG (in cv. Landsberg erecta).
 NLRD -> DLSYHE (in cv. Di-17).
 RVDTE -> SVNKK (in cv. Landsberg erecta).
 Q -> B (in cv. Di-17).
 YL -> HI (in cv. Di-17).
 Missing (in cv. Di-17).
 M -> K (in cv. Landsberg erecta).

Query Match
 Best Local Similarity 70.0%; Score 35; DB 1; Length 910;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Query 1 EBYVPXGXDY 10
 Db 883 EKLVPGEEDY 892

Query Match
 Best Local Similarity 60.0%; Pred. No. 16;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Query 1 EBYVPXGXDY 10
 Db 885 EKLVPGEEDY 894

RESULT 3
 RPH8 ARATH STANDARD; PRT; 910 AA.
 AC P59584;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Disease resistance protein RPH8A (RPH8 homolog A).
 GN RPH8A.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Rubryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledones; core eudicots; Rosidae;
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A., AND FUNCTION.
 RC STRAIN=cv. Landsberg erecta;
 RX MEDLINE=99010193; PubMed=9811794;
 RA McDowell J.M., Dhandaydham M., Long T.A., Aarts M.G.M., Goff S.,
 RA Holub E.B., Dangl J.L.;
 RT "Intragenic recombination and diversifying selection contribute to the
 evolution of downy mildew resistance at the RPP8 locus of
 Arabidopsis";
 RL Plant Cell 10:1861-1874(1998).
 CC -1- FUNCTION: Disease resistance protein. Resistance proteins guard
 the plant against pathogens that contain an appropriate avirulence
 protein via an indirect interaction with this avirulence protein.
 CC That triggers a defense system including the hypersensitive
 response, which restricts the pathogen growth. In contrast to
 CC RPP8, it does not specifically recognize the Emco5 avirulence
 CC protein from Pterosporea parasitica.
 CC -1- MISCELLANEOUS: In cv. Columbia and cv. Di-17, this protein is not
 CC present due to an unequal crossing over between the RPP8 and RPH8A
 CC genes that creates a unique RPP8 gene.
 CC -1- SIMILARITY: Belongs to the disease resistance NB-LRR family.
 CC -1- RPP8/HRT subfamily.
 CC -1- SIMILARITY: Contains 2 leucine-rich (LRR) repeats.
 CC -1- SIMILARITY: Contains 1 NB-ARC domain.
 CC -1- DATABASES: NAME=NB-LRRS;
 CC NOTS=Functional and comparative genomics of disease resistance gene
 CC homologs;
 CC WWW=http://nblrrs.ucdavis.edu".
 KW Plant defense; ATP-binding; Repeat; Leucine-rich repeat.
 FT DOMAIN 10 45
 FT DOMAIN 146 459
 FT REPEAT 602 625
 FT REPEAT 844 869
 FT NP BIND 192 199
 FT SEQUENCE 910 AA; 105263 MW; 5B18P65A19A12BR CRC64;

```

RESULT 4
CARB_FUSNN STANDARD; PRT; 1058 AA.
ID CARB_FUSNN
AC Q8RG86;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
GN carb OR F04042.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteriia; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21866394; PubMed=11889109;
RA Kapriel V., Anderson I., Ivanova N., Reznik G., Los T., Lykdis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman B., Bernal A.,
RA Larsen N., D'Souza M., Malinas T., Pusch G., Haselkorn R.,
RA Forstner M., Kyriades N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
CC -1 CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -1 COPACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -1 PATHWAY: Arginine biosynthesis; first step.
CC -1 SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -1 SIMILARITY: BELONGS TO THE CARB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE010554; AAL94625.1; ALT_INIT.
DR HAMAP; MF 01210; -; 1.
DR InterPro; IPR006275; CARB_L_glu.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR004362; MGS_Like.
DR Pfam; PF00289; CPase_L_chain; 2.
DR Pfam; PF02786; CPase_L_D2; 2.
DR Pfam; PF02787; CPase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPASAB.
DR TIGRfam; TIGR01369; CPaseL1; 1.
DR PROSITE; PS00866; CPASAB_1; 2.
DR PROSITE; PS00867; CPASAB_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1058
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).

```

```

FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1058 AA; 117451 MM; ED7037AF77C1E39F CRC64;

Query Match
Best Local Similarity 60.0%; Score 34; DB 1; Length 1058;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EWPXGXGYS 11
Db 190 EIVPNGLNTS 199

RESULT 5
Y1LK TYDVA STANDARD; PRT; 102 AA.
ID Y1LK TYDVA
AC Y1LK TYDVA
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE Hypothetical 11.2 kDa protein.
GN VI.
OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
OC Viruses; ssDNA viruses; Geminiviridae; Maadrevirus.
OC NCBI_TaxID=31599;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92188538; PubMed=1546458;
RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;
RT "The nucleotide sequence of the infectious cloned DNA component of
RT tobacco yellow dwarf virus reveals features of geminiviruses
RT infecting monocotyledonous plants."
RL Virology 187:633-642(1992).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M81103; AAA47947.1; -.
DR PIR; A42452; A42452.
DR InterPro; IPR00621; Gemini_mov.
DR Pfam; PF01708; Gemini_mov; 1.
KW Hypothetical protein.
SQ SEQUENCE 102 AA; 11178 MM; A40BCF1E0AF55867 CRC64;

Query Match
Best Local Similarity 60.0%; Score 34; DB 1; Length 102;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EWPXGXGYS 11
Db 7 QVVPNGLNTS 16

RESULT 6
HMPA VIBCH STANDARD; PRT; 394 AA.
ID HMPA VIBCH
AC Q9RMV3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Flavohemoprotein (Hemoglobin-like protein) (Flavohemoglobin).
GN HMP OR VCA0183.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OC NCBI_TaxID=666;

```

[1]
 PN SEQUENCE FROM N.A.
 RP STRAIN=EI Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10953301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.B., Read T.D., Tettelin H., Richardson D.,
 RA Rasmussen L.M., Vamathevan J., Bass S., Qin H., Dragoi I., Seligman P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
 RT *cholerae*.";
 RL Nature 406:477-483(2000).
 CC -1- DOMAIN: CONSISTS OF TWO DISTINCT DOMAINS; ONE IS A HEME-CONTAINING
 CC OXYGEN BINDING DOMAIN IN THE N-TERMINAL REGION AND THE OTHER IS AN
 CC PAD-CONTAINING REDUCTASE DOMAIN FOUND IN THE C-TERMINAL REGION.
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY. TWO-DOMAIN
 CC FLAVOHEMOPROTEIN SUBFAMILY.
 CC -1- SIMILARITY: TO A NUMBER OF PAD/NAD(P) FLAVOPROTEIN
 CC OXIDOREDUCTASES.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB004358; AAF96096.1; -
 DR PIR: F82491; F82491.
 DR HSPG: P39662; ICQX.
 DR TIGR: VCA0183; -
 DR InterPro: IPR001834; Cyt B5 reductase.
 DR InterPro: IPR001709; PPN_Cyt_reductase.
 DR InterPro: IPR000971; Globin.
 DR InterPro: IPR001433; Oxred_PAD/NAD(P).
 DR InterPro: IPR001221; Phe hydroxylase.
 DR Pfam: PF00970; PAD binding_6; 1.
 DR Pfam: PF00042; Globin; 1.
 DR Pfam: PF00175; NAD binding_1; 1.
 DR PRINTS: PR00371; PFNCR.
 DR PRINTS: PR00410; PFHIDRXLASE.
 DR PROSITE: PS01033; GLOBIN; 1.
 DR Oxioreductase; NADP; Heme; Flavoprotein; PAD; Iron transport;
 KM Oxygen transport; Transport; Complete proteome.
 FT METAL 53 53 IRON (HEME DISTAL LIGAND)
 FT METAL 85 85 IRON (HEME PROXIMAL LIGAND)
 FT NP BIND 268 273 NADP (RIBOSE PART) (BY SIMILARITY).
 FT SEQUENCE 394 AA; 44191 MW; DDA3490FAE2882A CRC64;
 Query Match 68.0%; Score 34; DB 1; Length 394;
 Best Local Similarity 66.7%; Pred. NO. 11;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 2 EWPXGXDY 10
 Db 194 EPTPEGSDY 202
 RESULT 7
 PAF_DROME STANDARD; PRT; 2778 AA.
 AC P55824; Q9V9T6; Q9Y0Z7;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable ubiquitin carboxyl-terminal hydrolase PAF (EC 3.1.2.15)
 DE (Ubiquitin thiolesterase PAF) (Ubiquitin-specific processing protease

DE PAF) (Deubiquitinating enzyme PAF) (Fat facets protein).
 GN PAF OR BCDNA: D022582 OR CG1945.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Rhytididae; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND TISSUE SPECIFICITY.
 RC TISSUE-Specific imaginal disk;
 RX MEDLINE=93202020; PubMed=1295747;
 RA Fleischer-Vize J.A., Rubin G.M., Lehmann R.;
 RT "The fat facets gene is required for *Drosophila* eye and embryo
 RT development.";
 RL Development 116:985-1000(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoekstra R.A., Galle R.F.,
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champe M., Pfeiffer B.D.,
 RA Abertl J.F., Agbayani A., An H.-O., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beckson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolhaver S.,
 RA Borokova D., Botchan M.R., Bouck J., Brodeur P., Brothier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Dwyer D.E., Dew I., Dietz S.M.,
 RA Dodson R.A., Dou P.L., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodde C., Gong F., Gorelli J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hovstlin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
 RA Jalali M., Kalish F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Laeke P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulin G., Milshina N.V., Mobarry C., Morris J., Mostrel A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murthy D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Weissman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP REVISIONS, AND ALTERNATIVE SPLICING.
 RC STRAIN=Berkely.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Miera S., Crosby M.A., Mungall C.U., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.B.,
 RA Smith C.D., Tupy J.L., Whitfield B.J., Bayraktaroglu L., Berman B.P.,
 RA Betencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.B.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 RN [4]
 RP SEQUENCE OF 1089-2778 FROM N.A. (ISOFORM 1).

RC STRAIN-Berkeley;
 RX MEDLINE=20196012; PubMed=10731138;
 RA Rubin G.M., Hong L., Brokstein P., Evans-Holm M., Frise E.,
 RA Stapleton M., Harvey D.A.;
 RT "A Drosophila complementary DNA resource.";
 RL Science 287:2222-2224(2000).
 CC -1- FUNCTION: REQUIRED FOR EYE AND EMBRYO DEVELOPMENT, AND PLAYS A
 CC ROLE IN COMPOUND EYE ASSEMBLY AND COGENESIS RESPECTIVELY. IN THE
 CC LARVAL EYE DISKS, CELLS OUTSIDE THE ASSEMBLING FACETS REQUIRE THIS
 CC PROTEIN FOR SHORT-RANGE CELL INTERACTIONS THAT PREVENT THE MYSTERY
 CC CELLS FROM BECOMING PHOTORECEPTORS. IT IS ALSO REQUIRED FOR
 CC NUCLEAR MIGRATION AND CELLULARIZATION IN EARLY EMBRYOGENESIS AND
 CC COULD PLAY A ROLE IN POLE CELL DETERMINATION, DEVELOPMENT OR
 CC FUNCTION.
 CC -1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 CC ubiquitin + a thiol.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Comment=Experimental confirmation may be lacking for some
 CC isoforms;
 CC Name=1;
 CC IsoId=P55824-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=P55824-2; Sequence=VSP_005270;
 CC Name=3;
 CC IsoId=P55824-3; Sequence=VSP_005269;
 CC -1- TISSUE SPECIFICITY: EYE DISKS AND OVARIES.
 CC -1- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
 CC -1- SIMILARITY: Belongs to peptidase family C19.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L04959; AAF01345.1; -;
 CC EMBL; L04958; AAF01346.1; -;
 CC EMBL; L04960; AAF01347.1; -;
 CC EMBL; L04960; AAF01348.1; -;
 CC EMBL; AE003779; AAF57198.1; -;
 CC EMBL; AE003779; AAF57198.1; -;
 CC EMBL; AE003779; AAF57198.1; -;
 CC EMBL; AF145677; AAD38652.1; -;
 CC MEROPS; C19.007; -;
 CC Flybase; FBgn0005632; faf.
 CC GO; GO:0005737; C:cytoplasm; IDA.
 CC GO; GO:0007349; P:cellularization; IMP.
 CC GO; GO:000795; P:embryonic morphogenesis; IMP.
 CC GO; GO:0008583; P:myelocyte cell fate differentiation (sensu Dr. .); IMP.
 CC GO; GO:0007097; P:nuclear migration; IMP.
 CC GO; GO:0006512; P:ubiquitin cycle; IGI.
 CC InterPro; IPR001394; UCH-2.
 CC Pfam; PF00443; UCH_1.
 CC PROSITE; PS00972; UCH_2_1; 1.
 CC PROSITE; PS00973; UCH_2_2; 1.
 CC PROSITE; PS00235; UCH_2_3; 1.
 CC Ubl conjugation pathway; Hydrolase; Thiol protease;
 CC Developmental protein; Vision; Alternative splicing.
 CC ACT_SITE 1677 1677
 CC ACT_SITE 1677 1677
 CC ACT_SITE 1978 1978
 CC ACT_SITE 1986 1986
 CC ACT_SITE 1986 1986
 CC VARSPLIC 2705 2778
 CC -----
 CC VARSPLIC 2742 2778
 CC -----
 CC CONFLICT 234 234
 CC CONFLICT 2725 2725
 CC SEQUENCE 2778 AA; 311139 MM; FFB90438BA53A02B CRC64;

Query Match 68.0%; Score 34; DB 1; Length 2778;
 Best Local Similarity 54.5%; Pred. No. 86;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 REVPRGXGXDYS 11
 DB 1394 EYVDPGQDYS 1404
 RESULT 8
 REV STVCZ
 ID REV STVCZ STANDARD; PRT; 124 AA.
 AC P17280;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DB REV protein (anti-repression transactivator protein) (ART/TRS).
 GN REV.
 OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OC NCBI_TaxId=11723;
 RX MEDLINE=90259077; PubMed=2188136;
 RA Huec T., Cheyner R., Meyers A., Roelants G., Main-Hobson S.;
 RT "Genetic organization of a chimpanzee lentivirus related to HIV-1.";
 RL Nature 345:356-359(1990).
 CC -1- FUNCTION: REV APPEARS TO ACT POST-TRANSCRIPTIONALLY TO RELIEVE
 CC NEGATIVE REPRESSION OF GAG AND ENV PRODUCTION.
 CC -1- SUBCELLULAR LOCATION: Nuclear; accumulates in the nucleol.
 CC -1- PTM: PHOSPHOPROTEIN WHOSE STATE OF PHOSPHORYLATION IS MEDIATED
 CC BY A SPECIFIC SERINE KINASE ACTIVITY PRESENT IN THE NUCLEUS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X52154; CA36405.1; -;
 CC PIR; S09988; VKLJST.
 CC DR HIV; X52154; REVSCPZ.
 CC InterPro; IPR000625; REV_protein.
 CC Pfam; PF00424; REV_1.
 CC Transcription regulation; AIDS; Phosphorylation; Nuclear protein.
 CC SEQUENCE 124 AA; 13701 MM; F5877D1BD65A7B2 CRC64;
 Query Match 66.0%; Score 33; DB 1; Length 124;
 Best Local Similarity 60.0%; Pred. No. 5.4;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 EYVPRGXGXDYS 11
 DB 107 EYVPRGXGXDYS 116
 RESULT 9
 TPX LTSTN
 ID TPX LTSTN STANDARD; PRT; 165 AA.
 AC Q92BC5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DB Probable thiol peroxidase (EC 1.11.1.-).
 GN TPX OR LIN1625.
 OS Listeria innocua.
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 OC NCBI_TaxId=1642;
 RX (1)
 RP SEQUENCE FROM N.A.

```

RC STRAIN-CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Baqero P., Frangul L., Buchrieser C., Rusnok C., Amend A.,
RA Chabot A., Chetouani F., Couve E., de Daruvar A., Deloux P.,
RA Entlan K.-D., Fahl H., Garcia-del Portillo F., Garrido P.,
RA Gaudier L., Goebel U., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kunz M., Kunz F., Kuzupat G.,
RA Madeno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordisk B., Rose M., Schueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.;
RT "Comparative genomes of Listeria species.";
RU Science 294:849-852(2001).
CC -1- FUNCTION: Has antioxidant activity. Could remove peroxides or
CC H(2)O(2) (By similarity).
CC -1- SIMILARITY: Belongs to the ahpc/tea family. Tpx subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, AL596169; CAC96856.1; -.
DR PIR, AH1635; AH1635.
DR Listeria; L1N01625; -.
DR HAMAP, MF_00269; -.
DR InterPro, IPR000866; Ahpc-TSA.
DR InterPro, IPR002065; Tpx.
DR Pfam, PF00578; Ahpc-TSA; 1.
DR PROSITE, PS01265; Tpx; 1.
KW Antioxidant; Oxidoreductase; Peroxidase; Complete proteome.
SQ SEQUENCE 165 AA; 18162 MW; 77705B7CDB8C8F4D CRC64;

Query Match 66.0%; Score 33; DB 1; Length 165;
Best Local Similarity 66.7%; Pred. No. 7.3;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 EVVPGXNDY 10
DB 144 EVVPGSDH 152

RESULT 10
TPX_LISMO STANDARD; PRT; 165 AA.
AC Q8Y6U8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable thiol peroxidase (EC 1.11.1.-).
GN TPX OR LMO1583.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_Taxid=1639;

RP SEQUENCE FROM N.A.
RC STRAIN-EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Baqero P., Frangul L., Buchrieser C., Rusnok C., Amend A.,
RA Chabot A., Chetouani F., Couve E., de Daruvar A., Deloux P.,
RA Entlan K.-D., Fahl H., Garcia-del Portillo F., Garrido P.,
RA Gaudier L., Goebel U., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kunz M., Kunz F., Kuzupat G.,
RA Madeno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordisk B., Rose M., Schueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.,
RA Remmel B., Rose M., Schueter T., Simoes N., Tierrez A.,

```

```

RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.;
RT "Comparative genomes of Listeria species.";
RU Science 294:849-852(2001).
CC -1- FUNCTION: Has antioxidant activity. Could remove peroxides or
CC H(2)O(2) (By similarity).
CC -1- SIMILARITY: Belongs to the ahpc/tea family. Tpx subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, AL591979; CAC99661.1; -.
DR PIR, AG1272; AG1272.
DR Listeria; LMO01583; -.
DR HAMAP, MF_00269; -.
DR InterPro, IPR000866; Ahpc-TSA.
DR InterPro, IPR002065; Tpx.
DR Pfam, PF00578; Ahpc-TSA; 1.
DR PROSITE, PS01265; Tpx; 1.
KW Antioxidant; Oxidoreductase; Peroxidase; Complete proteome.
SQ SEQUENCE 165 AA; 18133 MW; 77705B7CC46D424D CRC64;

Query Match 66.0%; Score 33; DB 1; Length 165;
Best Local Similarity 66.7%; Pred. No. 7.3;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 EVVPGXNDY 10
DB 144 EVVPGSDH 152

RESULT 11
YX14_CABEL STANDARD; PRT; 327 AA.
AC P34338;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C15H7.4 in chromosome III.
GN C15H7.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;

RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Anscough R., Anderson K., Baynes C., Berke M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Pavellio A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston P., Jones M., Kersey J., Kirtlen J., Lister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Riffen L., Roopra A., Saunders D., Showken R.,
RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer B., Steden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Spiro J.,
RA Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.";
RU Nature 368:32-38(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@ib-sib.ch).
CC -----
DR EMBL; Z22173; CAAB0126.1; -.
DR PIR; S40753; S40753.
DR WormPep; C15H7.4; CE00082.
KM Hypothetical protein.
SQ SEQUENCE 327 AA; 35566 MW; 716BC2BD2E9607E CRC64;

Query Match 66.0%; Score 33; DB 1; Length 327;
Best Local Similarity 63.6%; Pred. No. 15;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXDYS 11
Db 175 KEVVPNGDKS 185

RESULT 12
ACDM_RAT STANDARD; PRT; 421 AA.
ID ACDM_RAT STANDARD; PRT; 421 AA.
AC P08503;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Acyl-CoA dehydrogenase, medium-chain specific, mitochondrial precursor
DE (EC 1.3.99.3) (MCAAD).
GN ACDM.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=87280028; PubMed=3611054;
RA Masebura Y., Krus J.P., Ozasa H., Glaesberg R., Finocchiaro G.,
RA Ikeda Y., Mole J., Rosenberg L.E., Tanaka K.;
RT "Molecular cloning and nucleotide sequence of cDNA encoding the
RT entire precursor of rat liver medium chain acyl coenzyme A
RT dehydrogenase.";
RL J. Biol. Chem. 262:10104-10108(1987).
CC -1- FUNCTION: THIS ENZYME IS SPECIFIC FOR ACYL CHAIN LENGTHS OF 4 TO
CC 16.
CC -1- CATALYTIC ACTIVITY: Acyl-CoA + ETF = 2,3-dihydroacyl-CoA + reduced
CC ETF.
CC -1- COFACTOR: FAD.
CC -1- PATHWAY: Mitochondrial fatty acid beta-oxidation system; first
CC step.
CC -1- SUBUNIT: Homotetramer.
CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -1- MISCELLANEOUS: A NUMBER OF STRAIGHT-CHAIN ACYL-COA DEHYDROGENASES
CC OF DIFFERENT SUBSTRATE SPECIFICITIES ARE PRESENT IN MAMMALIAN
CC TISSUES.
CC -1- SIMILARITY: BELONGS TO THE ACYL-COA DEHYDROGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ibt-sib.ch/announce/
CC or send an email to license@ib-sib.ch).
CC -----
DR EMBL; J02791; AAA40670.1; -.
DR PIR; A28436; DERTCM.
DR HSSP; P11310; 1EGD.
DR InterPro; IPR006089; Acyl-CoA_dh.
DR InterPro; IPR006090; Acyl-CoA_dh_C.
DR InterPro; IPR006091; Acyl-CoA_dh_M.
DR InterPro; IPR006092; Acyl-CoA_dh_N.
DR Pfam; PF00441; Acyl-CoA_dh; 1.
DR Pfam; PF02770; Acyl-CoA_dh_M; 1.
DR Pfam; PF02771; Acyl-CoA_dh_N; 1.

```

```

DR PROSITE; PS00072; ACYL_COA_DH_1; 1.
DR PROSITE; PS00073; ACYL_COA_DH_2; 1.
KM Oxidoreductase; Flavoprotein; FAD; Fatty acid metabolism;
KM Mitochondrion; Transit peptide.
FT TRANSIT 1 25
FT CHAIN 26 421
FT ACT_SITE 193 193
FT ACT_SITE 401 401
FT ACT_SITE 401 401
SQ SEQUENCE 421 AA; 46555 MW; 2CF076FC919BDB8 CRC64;

Query Match 66.0%; Score 33; DB 1; Length 421;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXDY 10
Db 58 ERIIPVAPDY 67

RESULT 13
SYR_STRP3 STANDARD; PRT; 563 AA.
ID SYR_STRP3 STANDARD; PRT; 563 AA.
AC O8K5J2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Arginyl-L-tyrosine synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ArgRS).
GN ARG8 OR SPY3_1809 OR SP31807.
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=198466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=22133806; PubMed=12122206;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Mueser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SGI-1 / Serotype M3;
RA Nakagawa I., Kurokawa K., Nakata M., Tomiyasu Y., Yamashita A.,
RA Yamazaki K., Okahashi N., Kawabata S., Yasunaga T., Hattori M.,
RA Hayaishi H., Hamada S.;
RT "The genome of invasive Streptococcus pyogenes; a comparative analysis
RT of S. pyogenes SSI-1, SF370 and MGAS8232.";
RL Submitted (May-2002) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
CC dihydrophosphate + L-arginyl-tRNA(Arg).
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ibt-sib.ch/announce/
CC or send an email to license@ib-sib.ch).
CC -----
DR EMBL; AE014171; AAM00416.1; -.
DR EMBL; AP005146; BAC64902.1; -.
DR HAMAP; MF_00123; -.
DR InterPro; IPR001278; Arg_tRNA-synt_1c.

```

DR InterPro; IPR005148; N.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF03485; N-Arg; 1.
 DR Pfam; PF00750; tRNA-synt_1d; 1.
 DR PRINTS; PRO1038; TRNASYNTHARG.
 DR TIGRFAMs; TIGR00456; args; 1.
 DR PROSITE; PS00178; AA tRNA_LIGASE I; FALSE NEG.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 DR SITE 121 "HIGH" REGION.
 FT SEQUENCE 563 AA; 63134 MW; B0B7DEC31A9DC63 CRC64;
 Query Match 66.0%; Score 33; DB 1; Length 563;
 Best Local Similarity 45.5%; Pred. No. 26;
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 BEVVPXGXKDY 11
 DB 94 EQVITAGSDYA 104
 RESULT 14
 STR_STRP8 STANDARD; PRT; 563 AA.
 AC Q8NZ22;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ARGRS).
 GN ARGS OR SP2151.
 OS Streptococcus pyogenes (serotype M18).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus;
 OC NCBI_TaxID=186103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MGAS8232 / Serotype M18;
 RX MEDLINE=21927593; PubMed=11917108;
 RA Smoot J.C., Barblian K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
 RA Sylva G.L., Sturdevant D.B., Ricklefs S.M., Porcella S.F.,
 RA Partins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
 RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
 RT "Genome sequence and comparative microarray analysis of serotype M18
 RT group A Streptococcus strains associated with acute rheumatic fever
 RT outbreaks";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
 CC -1- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
 CC diphosphate + L-arginyl-tRNA(Arg).
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE010119; AL98627.1; -.
 DR HAMAP; MF_00123; -; 1.
 DR InterPro; IPR001278; Arg_tRNA-synt_1c.
 DR InterPro; IPR005148; N.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF03485; N-Arg; 1.
 DR Pfam; PF00750; tRNA-synt_1d; 1.
 DR PRINTS; PRO1038; TRNASYNTHARG.
 DR TIGRFAMs; TIGR00456; args; 1.
 DR PROSITE; PS00178; AA tRNA_LIGASE I; FALSE NEG.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 DR SITE 121 "HIGH" REGION.
 FT SEQUENCE 563 AA; 63134 MW; B0B7DEC31A9DC63 CRC64;

SEQ SEQUENCE 563 AA; 63134 MW; F8C8F176A68DBB5B CRC64;

Query Match 66.0%; Score 33; DB 1; Length 563;
 Best Local Similarity 45.5%; Pred. No. 26;
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXKDY 11
 DB 94 EQVITAGSDYA 104

RESULT 15

STR_STRP8 STANDARD; PRT; 563 AA.

AC Q99XL5;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ARGRS).

GN ARGS OR SP22151.

OS Streptococcus pyogenes.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus;

OC NCBI_TaxID=1314;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SP370 / ATCC 700294 / Serotype M1;

RX MEDLINE=21192684; PubMed=11296296;

RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,

RA Plameaux C., Sezate S., Suvorov A.N., Kanton S., Lai H.S., Lin S.P.,

RA Qian Y., Jia H.G., Najjar P.Z., Ren Q., Zhu H., Song L., White J.,

RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;

RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes";

RT Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).

CC -1- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +

CC diphosphate + L-arginyl-tRNA(Arg).

CC -1- SUBUNIT: Monomer (By similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AB006633; AAK34788.1; -.

DR HAMAP; MF_00123; -; 1.

DR InterPro; IPR001278; Arg_tRNA-synt_1c.

DR InterPro; IPR005148; N.

DR InterPro; IPR001412; tRNA-synt_1.

DR Pfam; PF03485; N-Arg; 1.

DR Pfam; PF00750; tRNA-synt_1d; 1.

DR PRINTS; PRO1038; TRNASYNTHARG.

DR TIGRFAMs; TIGR00456; args; 1.

DR PROSITE; PS00178; AA tRNA_LIGASE I; FALSE NEG.

KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

KW Complete proteome.

DR SITE 121 "HIGH" REGION.

FT SEQUENCE 563 AA; 63120 MW; B0F2CAC28D03B613 CRC64;

Query Match 66.0%; Score 33; DB 1; Length 563;
 Best Local Similarity 45.5%; Pred. No. 26;
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXKDY 11
 DB 94 EQVITAGSDYA 104

Search completed: December 22, 2003, 17:42:30
Job time : 5.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:27:26 / Search time 25.2 Seconds

(without alignments)
112.642 Million cell updates/sec

Title: US-09-909-164-46
Perfect score: 50
Sequence: 1 EEVVPXGDYS 11

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SPTREMBL_23.*

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_protist.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriophage.*
17: sp_archaeop.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	72.0	298	2	052367 rhizobium t
2	36	72.0	319	16	Q8E8Y7 ocaenobacti
3	36	72.0	363	17	Q30260 archaeoglob
4	36	72.0	595	10	Q94677 androgaphi
5	36	72.0	1044	16	Q8D1H0 rhizobium t
6	35	70.0	78	6	Q9A8T4 caulis fami
7	35	70.0	143	5	Q9V8Y8 diatrophila
8	35	70.0	175	6	Q8I033 bos taurus
9	35	70.0	215	6	Q8I031 bos taurus
10	35	70.0	217	4	Q00404 homo sapien
11	35	70.0	290	16	Q8U730 agrobacteri
12	35	70.0	297	6	Q8H8Y3 bos taurus
13	35	70.0	299	4	Q9UEB9 homo sapien
14	35	70.0	582	16	Q9J2P8 neisseria m
15	35	70.0	697	5	Q9V190 diatrophila
16	35	70.0	906	10	Q9Z8Y4 arabidopsis

17	35	70.0	908	10	Q8W4J9 arabidopsis
18	35	70.0	908	10	Q9Z8Y3 arabidopsis
19	35	70.0	908	10	Q9FJX8 arabidopsis
20	35	70.0	909	10	Q9M5A1 arabidopsis
21	34	68.0	156	3	Q12479 saccharomyc
22	34	68.0	175	10	Q8VY88 arabidopsis
23	34	68.0	247	16	Q92U66 rhizobium m
24	34	68.0	257	10	Q9C6J0 arabidopsis
25	34	68.0	283	16	Q8N816 arabidopsis
26	34	68.0	433	16	Q9A382 caulobacter
27	34	68.0	1442	17	Q9V7H5 sulfolobus
28	33	66.0	143	17	Q8T6X2 methanopyru
29	33	66.0	196	16	Q8ZAO9 yersinia pe
30	33	66.0	210	10	Q65890 cycloclotella
31	33	66.0	222	16	Q8E5L9 streptococc
32	33	66.0	222	16	Q8DZM9 streptococc
33	33	66.0	225	10	Q40129 lycopersico
34	33	66.0	253	16	Q8XP88 clostridium
35	33	66.0	283	16	Q8FPY9 cornebacte
36	33	66.0	312	11	Q9D876 mus musculu
37	33	66.0	368	2	Q8K1A3 rhizobium e
38	33	66.0	440	17	Q9Y7I3 aeropyrum p
39	33	66.0	471	11	Q8R126 mus musculu
40	33	66.0	484	11	Q8VD18 mus musculu
41	33	66.0	484	11	Q8B7X4 mus musculu
42	33	66.0	484	11	Q8BK35 mus musculu
43	33	66.0	517	16	Q8X2L5 talistonia s
44	33	66.0	563	13	Q9DDJ4 halichoeres
45	33	66.0	678	12	Q9E1X6 cercopithec

ALIGNMENTS

RESULT 1

ID	052367	PRELIMINARY;	PRT;	298 AA.
AC	052367	01-JUN-1998 (TREMBLrel. 06, Created)		
DT	01-JUN-1998	(TREMBLrel. 06, Last sequence update)		
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)		
DB	ARY1-alcohol dehydrogenase homolog (Fragment).			
GN	XYL1.			
OS	Rhizobium tropici.			
OC	Plasmid pRRCFN298.			
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;			
OC	Rhizobiaceae; Rhizobium.			
OX	NCBI_TaxID=398;			
RM	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-CFN298;			
RA	Rosenblueth M., Hynes M.F., Martinez-Romero E.;			
RL	Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.			
CC	-1- COFACTOR: ZINC (BY SIMILARITY).			
CC	-1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE			
CC	FAMILY.			
CC	EMBL; AF036920; AAC04779.1; -			
DR	InterPro; IPR002328; ADH_zinc.			
DR	InterPro; IPR002085; Adh_zn family.			
DR	InterPro; IPR00205; NAD binding.			
DR	Pfam; PF0107; adh_zinc.1.			
DR	PROSITE; PS00059; ADH_ZINC.1.			
KM	Metal-binding; Oxidoreductase; zinc; Plasmid.			
FT	NON_TER 298			
SO	SEQUENCE 298 AA; 31092 MW; 49B2P8117C3JAE87 CRC64;			
Query Match	72.0%; Score 36; DB 2; Length 298;			
Best local similarity	50.0%; Pred. NO. 15;			
Matches	5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;			
Qy	2 EEVVPXGDYS 11			
DB	250 EIIPEGADFS 259			

RESULT 2
Q8ESV7 PRELIMINARY; PRT; 319 AA.
AC Q8ESV7;
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Hypothetical conserved protein.
GN O80509.
OS Oceanobacillus thevenis.
OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.
OX NCBI_TaxID=162710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.,
RT "Genome sequence of Oceanobacillus thevenis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments."
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AP004594; BAC12465.1; -
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 319 AA; 35617 MW; 3BDAB8F13E79E37 CRC64;
QY 1 EEVVPXGXDY 10
Db 189 EQLVPHGIDY 198
Query Match 72.0%; Score 36; DB 16; Length 319;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
RESULT 3
Q30260 PRELIMINARY; PRT; 363 AA.
AC Q30260;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Hypothetical protein AF2411.
GN AF2411.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Kleck H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.B.,
RA Richardson D.L., Dodson R.J., Gilm M., Hickey E.K., Peterson J.D.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirsnes E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodex A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow F.W., D'Andrea K.F., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.,
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL; AF001109; AAB91255.1; -
DR TIGR; AF2411; -
DR InterPro; IPR002103; Bac_Juiciferase.
DR Pfam; PF00296; bac_juiciferase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 363 AA; 41736 MW; 0B976AE78F8F4803 CRC64;

Query Match 72.0%; Score 36; DB 17; Length 363;
Best Local Similarity 54.5%; Pred. No. 18;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 EEVVPXGXDY 11
Db 120 ENIVPYGIDFS 130
RESULT 4
Q946J7 PRELIMINARY; PRT; 595 AA.
AC Q946J7;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE 3-hydroxy-3-methylglutaryl-coenzyme A reductase (BC 1.1.1.34).
GN HMGR1.
OS Andropogon paniculata.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Lamiales; Rutaceae; Acanthaceae; Acanthaceae; Ruellieae;
OC Andropogonaceae; Andropogonaceae.
OX NCBI_TaxID=175694;
RN [1]
RP SEQUENCE FROM N.A.
RA Krishnan S., Banerjee N.S.,
RT "3-hydroxy-3-methylglutaryl coenzyme A reductase gene from
RT Andropogon paniculata.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF389879; AL28015.2; -
KM Oxidoreductase.
SQ SEQUENCE 595 AA; 63268 MW; 19A3BA572F67A82E CRC64;
QY 2 EEVVPXGXDY 11
Db 15 EVAPPGHDYS 24
Query Match 72.0%; Score 36; DB 10; Length 595;
Best Local Similarity 70.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
RESULT 5
Q8DIH0 PRELIMINARY; PRT; 1044 AA.
AC Q8DIH0;
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Multidrug efflux transporter.
GN TLL1618.
OS Synecchococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synecchococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Wakamabe A., Iriyuchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsumoto M., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.,
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005374; BAC09170.1; -
KM Complete proteome.
SQ SEQUENCE 1044 AA; 113205 MW; 00B9C13F0F63ED2F CRC64;
Query Match 72.0%; Score 36; DB 16; Length 1044;
Best Local Similarity 63.6%; Pred. No. 62;

Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGXDXS 11
 DB 843 EEVLPNGIGYS 853

RESULT 6

Q9XST4 PRELIMINARY; PRT; 78 AA.

AC Q9XST4; 01-NOV-1999 (TRENBLrel. 12, Created)

DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)

DT 01-DEC-2001 (TRENBLrel. 13, Last annotation update)

DE P97 homologous protein (Fragment).

CS P97.

OS Carls familiaris (Dog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

OC NCBI_Taxid=9615;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUR=Thyroid;

RX MEDLINE=20422104; PubMed=10964405;

RA Pichon B., Mercan D., Pouillon V., Christophe-Hobertus C.,

RA Christophe D.;

RT "A method for the large-scale cloning of nuclear proteins and nuclear

RT targeting sequences on a functional basis."

RL Anal. Biochem. 284:231-239(2000).

DR EMBL; AJ388531; CAB4683.1; -.

FT NON TER 78

FT SEQUENCE 78 AA; 8895 MW; B62486313555FBA1 CRC64;

Query Match 70.0%; Score 35; DB 6; Length 78;
 Best Local Similarity 54.5%; Pred. No. 5.3;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGXDXS 11
 DB 16 EDVVPGGGYS 26

RESULT 7

Q9VSY8 PRELIMINARY; PRT; 143 AA.

AC Q9VSY8; 01-MAY-2000 (TRENBLrel. 13, Created)

DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)

DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)

DE CG3911 protein.

GN CG3911.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OC NCBI_Taxid=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Gelinker S.B., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.B., Li P.W., Hoskins R.A., Galie R.P.,

RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,

RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Abuyant A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale U., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotcher P.,

RA Burris K.C., Busam D.A., Butler H., Cadieu B., Center A., Chandra I.,

RA Cherry J.M., Chewley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foaier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jaleli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun B.,

RA Svirskas R., Tector C., Turner R., Venter B., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers B.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195(2000).

DR EMBL; AB003552; AAP50270.1; -.

DR FLYBase; FBgn003592; CG3911.

DR SEQUENCE 143 AA; 16471 MW; 1DC346DC22C02AA2 CRC64;

Query Match 70.0%; Score 35; DB 5; Length 143;
 Best Local Similarity 66.7%; Pred. No. 11;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGXDX 9
 DB 135 EEVVPAGED 143

RESULT 8

Q8I033 PRELIMINARY; PRT; 175 AA.

AC Q8I033; 01-MAR-2003 (TRENBLrel. 23, Created)

DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)

DE Human-type bent protein.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.

OC NCBI_Taxid=9913;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Jersey;

RA Iwashita S., Itoh T.;

RT "A LINE-mediated gene diversity."

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB081003; BAC11952.1; -.

DR SEQUENCE 175 AA; 19529 MW; CE7283CB98393BB6 CRC64;

Query Match 70.0%; Score 35; DB 6; Length 175;
 Best Local Similarity 54.5%; Pred. No. 13;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGXDXS 11
 DB 16 EDVVPGGGYS 26

RESULT 9

Q8I031 PRELIMINARY; PRT; 215 AA.

AC Q8I031; 01-MAR-2003 (TRENBLrel. 23, Created)

DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)

DE Human-type bent protein.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.

OC NCBI_Taxid=9913;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Jersey;

RA Iwashita S., Itoh T.;

RT "A LINE-mediated gene diversity."

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB081003; BAC11952.1; -.

DR SEQUENCE 215 AA; 22529 MW; CE7283CB98393BB6 CRC64;

```

DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE H-type bcnt protein (Fragment).
GN H-TYPE BCNT.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RZPD #750;
RA Iwashita S., Itoh T., Sezaki M., Oshima K., Hashimoto E.,
RA Kitagawa Y., Takahashi T., Masui T., Hashimoto K.;
RT "A LINE-mediated gene diversity."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB081095; BAC15593.1; -.
FT NON TER 215 215
SQ SEQUENCE 215 AA; 23822 MW; 7287C8B98E9D05C8 CRC64;

Query Match 70.0%; Score 35; DB 6; Length 215;
Best Local Similarity 54.5%; Pred. No. 17;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXDXS 11
DB 16 EDYVPSGGGRYS 26

RESULT 10
ID 000404 PRELIMINARY; PRT; 217 AA.
AC 000404;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE P97 homologous protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RC MEDLINE=97160586; PubMed=9006920;
RA Nobukuni T., Kobayashi M., Comori A., Ichinose S., Iwanaga T.,
RA Takahashi I., Hashimoto K., Hattori S., Kabuchi K., Miyata Y.;
RT "An Ahb-linked repetitive sequence corresponding to 280 amino acids is
RT expressed in a novel bovine protein, but not in its human homologue."
RL J. Biol. Chem. 272:2801-2807(1997).
DR EMBL; D85939; BAA20069.1; -.
SQ SEQUENCE 217 AA; 24061 MW; E404BA2E35497828 CRC64;

Query Match 70.0%; Score 35; DB 4; Length 217;
Best Local Similarity 54.5%; Pred. No. 17;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXDXS 11
DB 16 EDYVPSGGGRYS 26

RESULT 11
ID 08U7J0 PRELIMINARY; PRT; 290 AA.
AC 08U7J0;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE 6-O-methylguanine-DNA methyltransferase.
GN ADA OR ATU4459 OR AGR L 816.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).

```

```

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OC NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monke D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.P. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eilen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Decherage G., Gillet M., Grant C.,
RA Kutyavina T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-P., Gordon M.P., Olson M.V.,
RA Neeter B.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Oucello B., Goldman B.S., Cao Y., Akenazi M., Halling C., Mullin L.,
RA Houwiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu P.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Planagan C., Crowell C., Gerson J., Lomo C., Sear C., Strub G.,
RA Clelo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL; AB009374; AL445253.1; -.
DR EMBL; AB008240; AAK8982.1; -.
DR InterPro; IPR000005; HTHARAC.
DR InterPro; IPR001497; Methyltransf_1.
DR Pfam; PF00165; HTH_Arac; 2.
DR Pfam; PF01035; Methyltransf_1; 1.
DR SMART; SM00342; HTH_ARAC; 1.
DR TIGRFAMs; TIGR00589; Ogt; 1.
DR PROSITE; PS01124; HTH_ARAC_FAMILY_2; 1.
KV Methyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 290 AA; 31587 MW; B6265928F519977F CRC64;

Query Match 70.0%; Score 35; DB 16; Length 290;
Best Local Similarity 50.0%; Pred. No. 24;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXDX 10
DB 9 EDITPISGDY 18

RESULT 12
ID 08HX9 PRELIMINARY; PRT; 297 AA.
AC 08HX9;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Human-type Bcnt.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Iwashita S., Itoh T.;
RT "A LINE-mediated gene diversity."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB081004; BAC11953.1; -.
SQ SEQUENCE 297 AA; 33354 MW; D4A944B6C740373C CRC64;

```

```

Query Match          70.0%; Score 35; DB 6; Length 297;
Best Local Similarity 54.5%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EVVVPXGXDYS 11
Db 16 EDVVPSCGEYS 26

RESULT 13
Q9UEB9 PRELIMINARY; PRT; 299 AA.
AC Q9UEB9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE BCTNT protein (CRANIOPACIAL development protein 1).
GN BCTNT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98267221; PubMed=9602175;
RA Takahashi I., Nobukuni T., Ohmori H., Kobayashi M., Tanaka S.,
RA Ohshima K., Okada N., Masui T., Hashimoto K., Iwashita S.,
RT "Existence of a bovine LINE repetitive insert that appears in the CDNA
RT of bovine protein BCTNT in ruminant, but not in human, genomes."
RL Gene 211:387-394(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (NCV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB009285; BAA31867.1; -
DR EMBL; BC000991; AA000991.1; -
DR Genew; HGNC:1873; CPDPI.
SQ SEQUENCE 299 AA; 33593 MW; F4A9B928B669451A CRC64;

Query Match          70.0%; Score 35; DB 4; Length 299;
Best Local Similarity 54.5%; Pred. No. 25;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EVVVPXGXDYS 11
Db 16 EDVVPSCGEYS 26

RESULT 14
Q9JZP8 PRELIMINARY; PRT; 587 AA.
ID Q9JZP8;
AC Q9JZP8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Succinate dehydrogenase, flavoprotein subunit.
GN NMB0950.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_Taxid=491;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MC58 / serogroup B;
RC MEDLINE=20175755; PubMed=10710307;
RA Teteijn H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eiden J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Macon T., Glazko A., Parksey D.S., Blair E., Cline H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathavan J.,

```

```

RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Frazer C.M., Moxon E.R., Rappunli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
CC -1- COFACTOR: PAD (BY SIMILARITY).
DR EMBL; AB002446; AAF41356.1; -
DR HSSP; P00363; 1KF6.
DR TIGR; NMB0950; -.
DR InterPro; IPR003953; PAD_bind2.
DR InterPro; IPR001327; PAD_Pyr_redox.
DR InterPro; IPR003952; FRD/SDH_PAD.
DR InterPro; IPR001100; Pyr_redox.
DR InterPro; IPR004112; Succ_DH_flav_C.
DR Pfam; PF00890; PAD_binding_2_1.
DR Pfam; PF02910; succ_DH_flav_C_1.
DR PRINTS; PR00368; PADPNT.
DR PRINTS; PR00411; PNDRTYASR1.
DR PROSITE; PS00504; FRD_SDH_PAD_BINDING; 1.
KW PAD; Flavoprotein; Oxidoreductase; Complete proteome.
SQ SEQUENCE 587 AA; 64502 MW; 9581701B08069003 CRC64;

Query Match          70.0%; Score 35; DB 16; Length 587;
Best Local Similarity 70.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EVVVPXGXDY 10
Db 366 EVVVPQGEYD 375

RESULT 15
Q9Y190 PRELIMINARY; PRT; 692 AA.
ID Q9Y190;
AC Q9Y190;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE CG10040 protein (Roughened eye)
GN RN OR CG10040 OR CG14600 OR CG14601 OR CG14603 OR CG12466.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailes R.M., Baxevanis A., Baxendale J., Bayraktaroglu L., Beasley B.M.,
RA Beeson K.Y., Bens P.V., Bernan B.P., Bhandari D., Bolintsov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E.J., Canter A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos R., Delcher A., Deng Z., Maye A.D., Dew I., Dietz S.M.,
RA Durbin K.J., Evans G.L., Fagan M., Garg N.S., Gelbart W.M., Glasser K.,
RA Foster C., Gabrielian A.B., Gary N.S., Gault J., Glavet S., Fleischmann W.,
RA Glodok A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Houten D., Houshun K.A., Howland T.J., Wei M.-H., Ileguam C.,
RA Jostali M., Kalush F., Karpen G.H., Ke Z., Kenton J.A., Ketchum K.A.,
RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,

```

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
 RA Palazzolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Glibbs R.A., Myers B.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).

RM [2]
 RP SEQUENCE FROM N.A.

RA St Pierre S.E., Galindo M.I., Couse J.P., Thor S.;

RT "Control of *Drosophila* imaginal disc development by rotund and
 RT roughened eye: differentially expressed transcripts of the same gene

RL encoding distinct zinc finger proteins.";
 RT Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL; AF003672; AAF54035.1; -

DR EMBL; AF395904; AAL59598.1; -

DR HSSP; P08153; 1ZFD.

DR FlyBase; FBgn0003263; m.

DR InterPro; IPR007087; znf_C2H2.

DR Pfam; PF00096; zf-C2H2; 5.

DR ProDom; PD000003; znf_C2H2; 1.

DR SMART; SM00355; znf_C2H2; 5.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.

DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 5.

KW Metal-binding; zinc; zinc-finger.
 SQ SEQUENCE 692 AA; 75774 MW; BD010502BB650428 CRC64;

Query Match 70.0%; Score 35; DB 5; length 692;
 Best Local Similarity 60.0%; Pred. NO. 65;

Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVFXGDYS 11

Db 26 EIPFAGDYS 35

Search completed: December 22, 2003, 17:51:33
 Job time : 25.2667 secs

PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
DR
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
XX virus protease -
XX
PS Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EEVVPXGMSYS 11
DB 1 EEVVPXGMSYS 11
RESULT 2
ABB80522
ID ABB80522 standard; peptide; 11 AA.
XX
AC ABB80522;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.
XX
KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C

PT virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
PS
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EEVVPXGMSYS 11
DB 1 EEVVPXGMSYS 11
RESULT 3
ABB80525
ID ABB80525 standard; peptide; 11 AA.
XX
AC ABB80525;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #5.
XX
KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
XX virus protease -
XX
PS Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of

CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX

Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHVVPXGMSYS 11
DB 1 EHVVPXGMSYS 11

RESULT 4
ABB80526
ID ABB80526 standard; peptide; 11 AA.

AC ABB80526;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #6.

KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketamide;
virocide.

OS Synthetic.

Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

PN W0200208251-A2.

XX 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

PA (CORV-) CORVAS INT INC.

PI Llm-wlby M, Levy OB, Brunck TK;

PS WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -

PS Claim 17; Page 64; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX

Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHVVPXGMSYS 11
DB 1 EHVVPXGMSYS 11

RESULT 5
ABB80559
ID ABB80559 standard; peptide; 11 AA.

AC ABB80559;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #39.

KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketamide;
virocide.

OS Synthetic.

Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Modified-site 8 /note= "Oxymethionine"

FT Modified-site 11 /note= "C-terminal amide"

PN W0200208251-A2.

XX 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

PA (CORV-) CORVAS INT INC.

PI Llm-wlby M, Levy OB, Brunck TK;

PS WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -

PS Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX

Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHVVPXGMSYS 11
|||||


```

FT      Modified-site      /note= "N-terminal acetyl"
PT      6
FT      /note= "Norleucyl carbonyl forming keto-amide linkage
FT      with residue 7"
FT      Modified-site      11
FT      /note= "C-terminal amide"
PN      WO200208251-A2.
XX      31-JAN-2002.
XX      19-JUL-2001; 2001WO-US23169.
XX      21-JUL-2000; 2000US-220101P.
XX      (CORV-) CORVAS INT INC.
XX      Lim-wilby M, Levy OE, Brunck TK;
XX      WPI; 2002-361643/39.
XX      Novel peptide compound having hepatitis C virus protease inhibitory
PT      activity useful for treating disorders associated with hepatitis C
XX      virus protease
XX      Claim 17; Page 65; 69pp; English.
XX      The sequence represents a peptide compound of the invention having
CC      hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC      invention are alpha-ketoamide peptide analogues. The peptides have
CC      virucide activity, and are useful for treating and in the manufacture of
CC      a medicament to treat disorders associated with HCV protease. A
CC      pharmaceutical composition comprising the peptide as an active ingredient
CC      is useful for treating disorders associated with hepatitis C virus.
XX      Sequence 11 AA;
SQ
XX
XX      Query Match      96.2%; Score 50; DB 23; Length 11;
XX      Best Local Similarity 100.0%; Pred. No. 0.0013;
XX      Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 EKVVPXGMSYS 11
DB      1 EKVVPXGMSYS 11
XX
XX      RESULT 9
XX      ABB80566
XX      ID ABB80566 standard; peptide; 11 AA.
XX      AC ABB80566;
XX      08-OCT-2002 (first entry)
XX      Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #46.
XX      Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX      virucide.
XX      Synthetic.
XX      Key      Location/Qualifiers
XX      Modified-site 1 /note= "N-terminal acetyl"
XX      Modified-site 6 /note= "2-aminoisobutyryl carbonyl residue forming a
XX      Modified-site 11 /note= "C-terminal amide"
XX      Modified-site 11 /note= "C-terminal amide"
XX      WO200208251-A2.
XX      31-JAN-2002.

```

```

XX      19-JUL-2001; 2001WO-US23169.
XX      21-JUL-2000; 2000US-220101P.
XX      (CORV-) CORVAS INT INC.
XX      Lim-wilby M, Levy OE, Brunck TK;
XX      WPI; 2002-361643/39.
XX      Novel peptide compound having hepatitis C virus protease inhibitory
PT      activity useful for treating disorders associated with hepatitis C
XX      virus protease
XX      Claim 17; Page 65; 69pp; English.
XX      The sequence represents a peptide compound of the invention having
CC      hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC      invention are alpha-ketoamide peptide analogues. The peptides have
CC      virucide activity, and are useful for treating and in the manufacture of
CC      a medicament to treat disorders associated with HCV protease. A
CC      pharmaceutical composition comprising the peptide as an active ingredient
CC      is useful for treating disorders associated with hepatitis C virus.
XX      Sequence 11 AA;
SQ
XX
XX      Query Match      96.2%; Score 50; DB 23; Length 11;
XX      Best Local Similarity 100.0%; Pred. No. 0.0013;
XX      Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 EKVVPXGMSYS 11
DB      1 EKVVPXGMSYS 11
XX
XX      RESULT 10
XX      ABB80567
XX      ID ABB80567 standard; peptide; 11 AA.
XX      AC ABB80567;
XX      08-OCT-2002 (first entry)
XX      Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #47.
XX      Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX      virucide.
XX      Synthetic.
XX      Key      Location/Qualifiers
XX      Modified-site 1 /note= "N-terminal acetyl"
XX      Modified-site 6 /note= "(S,S)allothreonyl carbonyl residue forming a
XX      Modified-site 11 /note= "C-terminal amide"
XX      Modified-site 11 /note= "C-terminal amide"
XX      WO200208251-A2.
XX      31-JAN-2002.
XX      19-JUL-2001; 2001WO-US23169.
XX      21-JUL-2000; 2000US-220101P.
XX      (CORV-) CORVAS INT INC.
XX      Lim-wilby M, Levy OE, Brunck TK;
XX      WPI; 2002-361643/39.

```

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EHVVPXGMSYS 11
DB 1 EHVVPXGMSYS 11
RESULT 11
ABB80568
ID ABB80568 standard; peptide; 11 AA.
XX
XX ABB80568;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #48.
DE
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM virucide.
KM
XX Synthetic.
OS
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Alpha-propionyl-glycyl-carbonyl residue forming
FT a keto-amide linkage with residue 7"
FT Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EHVVPXGMSYS 11
DB 1 EHVVPXGMSYS 11
RESULT 12
ABB80524
ID ABB80524 standard; peptide; 11 AA.
XX
XX ABB80524;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.
DE
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM virucide.
KM
XX Synthetic.
OS
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 88.5%; Score 46; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.0088;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 1 BEVVPXGMDYS 11

RESULT 13

ABB80528
ID ABB80528 standard; peptide; 11 AA.

AC ABB80528;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #8.

KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketamide;
virocide.

OS Synthetic.

FT Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

FN WO200208251-A2.

PD 31-JAN-2002.

PE 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

PA (CORV-) CORVAS INT INC.

PI Lim-wilby M, Levy OB, Brunck TK;

DR WPI; 2002-361643/39.

PT Novel peptide compound having hepatitis C virus protease inhibitory
activity useful for treating disorders associated with hepatitis C
virus protease

PS Claim 17; Page 64; 69pp; English.

CC The sequence represents a peptide compound of the invention having
hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
invention are alpha-ketamide peptide analogues. The peptides have
virocide activity, and are useful for treating and in the manufacture of
a medicament to treat disorders associated with HCV protease. A
pharmaceutical composition comprising the peptide as an active ingredient
is useful for treating disorders associated with hepatitis C virus.

SO Sequence 11 AA;

QY Query Match 88.5%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0088;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 1 BEVVPXGMDYS 11

RESULT 14

ABB80529
ID ABB80529 standard; peptide; 11 AA.

AC ABB80529;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.

KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketamide;
virocide.

OS Synthetic.

FT Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

FN WO200208251-A2.

PD 31-JAN-2002.

PE 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

PA (CORV-) CORVAS INT INC.

PI Lim-wilby M, Levy OB, Brunck TK;

DR WPI; 2002-361643/39.

PT Novel peptide compound having hepatitis C virus protease inhibitory
activity useful for treating disorders associated with hepatitis C
virus protease

PS Claim 17; Page 64; 69pp; English.

CC The sequence represents a peptide compound of the invention having
hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
invention are alpha-ketamide peptide analogues. The peptides have
virocide activity, and are useful for treating and in the manufacture of
a medicament to treat disorders associated with HCV protease. A
pharmaceutical composition comprising the peptide as an active ingredient
is useful for treating disorders associated with hepatitis C virus.

SO Sequence 11 AA;

QY Query Match 88.5%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0088;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 1 BEVVPXGMDYS 11

RESULT 15

ABB80561
ID ABB80561 standard; peptide; 11 AA.

AC ABB80561;

XX

DT 08-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #41.
 XX
 KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1
 FT /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Modified-site 8
 FT /note= "Oxymethionine"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX
 FT W0200208251-A2.
 XX
 PN
 XX
 PD 31-JAN-2002.
 XX
 PP 19-JUL-2001; 2001WO-US23169.
 XX
 PR 21-JUL-2000; 2000US-220101P.
 XX
 PA (CORV-) CORVAS INT INC.
 XX
 PI Lin-wilby M, Levy OE, Brunck TK;
 DR WPI; 2002-361643/39.
 XX
 DR
 XX
 PT Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX
 XX
 PS Claim 17; Page 65; 69pp; English.
 XX
 CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a pharmaceutical to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 CC
 SO Sequence 11 AA;
 Query Match 88.5%; Score 46; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred.No. 0.0088;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EEVVPXGMSYS 11
 Db 1 EEVVPXGMDYS 11

Search completed: December 22, 2003, 17:41:03
 Job time : 33.4667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:37:03 ; Search time 10.133 Seconds
(without alignment)

45.930 Million cell updates/sec

Title: US-09-909-164-47
Perfect score: 52
Sequence: 1 BEVVPXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/aa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	65.4	947	4	US-09-228-986-73
2	34	65.4	1407	4	US-09-328-352-7885
3	33	63.5	45	2	US-08-637-759B-236
4	33	63.5	45	3	US-08-871-355A-236
5	33	63.5	45	4	US-09-201-945-236
6	33	63.5	65	6	5177197-51
7	33	63.5	410	6	5177197-1
8	33	63.5	1394	6	5177197-30
9	32	61.5	10	3	US-09-357-952-66
10	32	61.5	10	4	US-09-521-650-66
11	32	61.5	10	4	US-09-168-888-66
12	32	61.5	102	4	US-08-580-988A-23
13	32	61.5	152	2	US-08-460-694-4
14	32	61.5	152	3	US-08-460-744-4
15	32	61.5	152	3	US-07-667-711B-4
16	32	61.5	173	1	US-08-193-977-7
17	32	61.5	189	2	US-08-464-517-21
18	32	61.5	189	2	US-08-246-361A-21
19	32	61.5	189	2	US-08-463-772-21
20	32	61.5	189	5	PCT-US93-05000-21
21	32	61.5	236	2	US-08-464-517-22
22	32	61.5	236	2	US-08-246-361A-22
23	32	61.5	236	2	US-08-463-772-22
24	32	61.5	236	5	PCT-US93-05000-22
25	32	61.5	280	3	US-08-464-517-6
26	32	61.5	280	3	US-08-463-772-6
27	32	61.5	289	2	US-08-246-361A-4

28	32	61.5	289	5	PCT-US93-05000-4	Sequence 4, Appli
29	32	61.5	291	5	PCT-US93-05000-6	Sequence 6, Appli
30	32	61.5	292	2	US-08-464-517-23	Sequence 23, Appli
31	32	61.5	292	2	US-08-246-361A-23	Sequence 23, Appli
32	32	61.5	292	2	US-08-463-772-23	Sequence 23, Appli
33	32	61.5	292	3	US-08-463-772-23	Sequence 23, Appli
34	32	61.5	292	5	PCT-US93-05000-23	Sequence 23, Appli
35	32	61.5	295	1	US-07-947-120-8	Sequence 8, Appli
36	32	61.5	295	1	US-08-472-893A-8	Sequence 8, Appli
37	32	61.5	295	2	US-08-460-694-2	Sequence 2, Appli
38	32	61.5	295	2	US-08-464-517-19	Sequence 19, Appli
39	32	61.5	295	2	US-08-464-517-20	Sequence 20, Appli
40	32	61.5	295	2	US-08-246-361A-19	Sequence 19, Appli
41	32	61.5	295	2	US-08-246-361A-20	Sequence 20, Appli
42	32	61.5	295	3	US-08-463-772-19	Sequence 19, Appli
43	32	61.5	295	3	US-08-463-772-20	Sequence 2, Appli
44	32	61.5	295	3	US-08-460-744-2	Sequence 2, Appli
45	32	61.5	295	3	US-07-667-711B-2	Sequence 2, Appli

ALIGNMENTS

```

RESULT 1
US-09-228-986-73
Sequence 73, Application US/09228986
Patent No. 6359198
GENERAL INFORMATION:
APPLICANT: Strabala, Timothy
TITLE OF INVENTION: Compositions Isolated from Plant Cells
TITLE OR INVENTION: and Their Use in the Modification of Plant Cell Signalling
FILE REFERENCE: 11000/1020
CURRENT APPLICATION NUMBER: US/09/228,986
CURRENT FILING DATE: 1999-01-12
NUMBER OF SEQ ID NOS: 130
SOFTWARE: PasteSeq for Windows Version 3.0
SEQ ID NO 73
LENGTH: 947
TYPE: PRT
ORGANISM: Pinus radiata
US-09-228-986-73

Query Match
Best Local Similarity 65.4%; Score 34; DB 4; Length 947;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Cy 3 VVPXGMSYS 11
Db 686 VWPXGMSYS 694

RESULT 2
US-09-328-352-7885
Sequence 7885, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GPC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 7885
LENGTH: 1407
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-7885

Query Match
Best Local Similarity 65.4%; Score 34; DB 4; Length 1407;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

QY 2 EVVFXGMSY 10
|||:|:
Db 596 EVVPEGLSP 604

RESULT 3

US-08-637-759B-236
Sequence 236, Application US/08637759B
Patent No. 5876931
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/637,759B
FILING DATE: 03-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 236:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-637-759B-236

Query Match 63.5%; Score 33; DB 2; Length 45;
Best Local Similarity 60.0%; Pred. No. 6.6;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
||:|:|
Db 1 ERIISPLGMSY 10

RESULT 4

US-08-871-355A-236
Sequence 236, Application US/08871355A
Patent No. 6015669
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/871,355A
FILING DATE: 09-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 236:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-871-355A-236

CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/871,355A

FILING DATE: 09-JUN-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB95/02875

FILING DATE: 11-DEC-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Pabst, Patrea L.

REGISTRATION NUMBER: 31,284

REFERENCE/DOCKET NUMBER: RPMS 101 CON

TELECOMMUNICATION INFORMATION:

TELEPHONE: (404) 873-8794

TELEFAX: (404) 873-8795

INFORMATION FOR SEQ ID NO: 236:

SEQUENCE CHARACTERISTICS:

LENGTH: 45 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO
US-08-871-355A-236

Query Match 63.5%; Score 33; DB 3; Length 45;
Best Local Similarity 60.0%; Pred. No. 6.6;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
||:|:|
Db 1 ERIISPLGMSY 10

RESULT 5

US-09-201-945-236
Sequence 236, Application US/09201945
Patent No. 6342215
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/201,945
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/637,759
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:

```
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 236:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
US-09-201-945-236

Query Match      63.5%; Score 33; DB 4; Length 45;
Best Local Similarity 60.0%; Pred. No. 6.6;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      1 BEVVPXGMSY 10
       :|:|:|:|:|:|
       1 EELSLGMSY 10

Db

RESULT 6
517197-51
; Patent No. 517197
; APPLICANT: KANZAKI, TETSUO; OLOFSSON, ANDERS; MOREN, ANITA;
; WENSTEDT, CHRISTER; HELLMAN, ULF; MIYAZONO, KOHEI; CLAESSON-WELSH,
; LENA; HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETA1-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO: 51
; LENGTH: 65
517197-51

Query Match      63.5%; Score 33; DB 6; Length 65;
Best Local Similarity 45.5%; Pred. No. 10;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      1 BEVVPXGMSY 11
       :|:|:|:|:|:|
       52 KEICPGMGYT 62

Db

RESULT 7
517197-1
; Patent No. 517197
; APPLICANT: KANZAKI, TETSUO; OLOFSSON, ANDERS; MOREN, ANITA;
; WENSTEDT, CHRISTER; HELLMAN, ULF; MIYAZONO, KOHEI; CLAESSON-WELSH,
; LENA; HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETA1-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO: 1
; LENGTH: 410
517197-1

Query Match      63.5%; Score 33; DB 6; Length 410;
Best Local Similarity 45.5%; Pred. No. 81;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      1 BEVVPXGMSY 11
       :|:|:|:|:|:|
       399 KEICPGMGYT 409

Db
```

```
RESULT 8
517197-30
; Patent No. 517197
; APPLICANT: KANZAKI, TETSUO; OLOFSSON, ANDERS; MOREN, ANITA;
; WENSTEDT, CHRISTER; HELLMAN, ULF; MIYAZONO, KOHEI; CLAESSON-WELSH,
; LENA; HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETA1-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO: 30
; LENGTH: 1394
517197-30

Query Match      63.5%; Score 33; DB 6; Length 1394;
Best Local Similarity 45.5%; Pred. No. 3.3e+02;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      1 BEVVPXGMSY 11
       :|:|:|:|:|:|
       399 KEICPGMGYT 409

Db

RESULT 9
US-09-357-952-66
; Sequence 66, Application US/09357952
; Patent No. 6248904
; GENERAL INFORMATION:
; APPLICANT: Zhang, Han-Zhong
; APPLICANT: Cai, Sui Xiong
; APPLICANT: Drewe, John A.
; APPLICANT: Yang, Wu
; TITLE OF INVENTION: No. 6248904el Fluorescence Dyes and Their Applications for Whol
; TITLE OF INVENTION: Fluorescence Screening Assays for Caspases, Peptidases, Proteas
; FILS REFERENCE: 1735,0030001
; CURRENT APPLICATION NUMBER: US/09/357,952
; EARLIER FILING DATE: 1999-07-21
; EARLIER APPLICATION NUMBER: US 60/093,642
; EARLIER FILING DATE: 21-JUL-1998
; NUMBER OF SEQ ID NOS: 139
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 66
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURES:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-357-952-66

Query Match      61.5%; Score 32; DB 3; Length 10;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      1 BEVVPXGMSY 10
       :|:|:|:|:|:|
       1 DDIVPCMSY 10

Db

RESULT 10
US-09-521-650-66
; Sequence 66, Application US/09521650
; Patent No. 6335429
; GENERAL INFORMATION:
; APPLICANT: Weber, Eckard
; APPLICANT: Cai, Sui Xiong
; APPLICANT: Keana, John F.W.
; APPLICANT: Drewe, John A.
```

APPLICANT: Zhang, Han-Zhong
TITLE OF INVENTION: No. 633542921 Fluorogenic or Fluorescent Reporter Molecules and
TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence
TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the
TITLE OF INVENTION: Use Thereof
FILE REFERENCE: 1735.0290002
CURRENT APPLICATION NUMBER: US/09/521.650
CURRENT FILING DATE: 2000-03-08
EARLIER APPLICATION NUMBER: 09/168,888
EARLIER FILING DATE: 1998-10-09
EARLIER APPLICATION NUMBER: US 60/061,582
EARLIER FILING DATE: 1997-10-10
EARLIER APPLICATION NUMBER: US 09/033,661
EARLIER FILING DATE: 1998-03-03
NUMBER OF SEQ ID NOS: 142
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 66
LENGTH: 10
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-521-650-66

Query Match 61.5%; Score 32; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXMSY 10
Db 1 DDIVPCMSY 10

RESULT 11
US-09-168-888-66
Sequence 66, Application US/09168888
Patent No. 6342611
GENERAL INFORMATION:
APPLICANT: Weber, Eckard
APPLICANT: Cai, Sui Xiong
APPLICANT: Keana, John F.W.
APPLICANT: Drewe, John A.
APPLICANT: Zhang, Han-Zhong
TITLE OF INVENTION: No. 6342611 Fluorogenic or Fluorescent Reporter Molecules and
TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence
TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the
TITLE OF INVENTION: Use Thereof
FILE REFERENCE: 1735.0290002
CURRENT APPLICATION NUMBER: US/09/168,888
CURRENT FILING DATE: 1998-10-09
EARLIER APPLICATION NUMBER: US 60/061,582
EARLIER FILING DATE: 1997-10-10
EARLIER APPLICATION NUMBER: US 09/033,661
EARLIER FILING DATE: 1998-03-03
NUMBER OF SEQ ID NOS: 142
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 66
LENGTH: 10
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-168-888-66

Query Match 61.5%; Score 32; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXMSY 10
Db 1 DDIVPCMSY 10

RESULT 12
US-08-580-988A-23
Sequence 23, Application US/08580988A
Patent No. 5856161
GENERAL INFORMATION:
APPLICANT: Aggarwal et al.
TITLE OF INVENTION: Tumor Necrosis Factor
TITLE OF INVENTION: Receptor-I-Associated Protein Kinase And Methods
TITLE OF INVENTION: For Its Use
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSER: Dr. Benjamin A. Adler
STREET: 8011 Candle Lane
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77071
COMPUTER READABLE FORM:
MEDIUM TYPE: 1.44 Mb floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word for Macintosh
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/580,988A
FILING DATE: January 3, 1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benjamin Aaron Adler, Ph.D., J.D.
REGISTRATION NUMBER: 35,423
REFERENCE/DOCKET NUMBER: D5721CIP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-777-2321
TELEFAX: 713-777-6908
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 102 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: no
HYPOTHETICAL: no
ANTI-SENSE: no
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
US-08-580-988A-23

Query Match 61.5%; Score 32; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXMSY 10
Db 24 BEVVPXMSY 33

RESULT 13
US-08-460-694-4
Sequence 4, Application US/08460694
Patent No. 5856655
GENERAL INFORMATION:
APPLICANT: Arnold, Andrew
TITLE OF INVENTION: PRAD1 Cyclin and its cDNA
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSER: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington

```
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,694
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McConathy, Evelyn H.
REGISTRATION NUMBER: 35,279
REFERENCE/DOCKET NUMBER: 0609,4070002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 152 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-460-694-4

Query Match
Best Local Similarity 61.5%; Score 32; DB 2; Length 152;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Cy 1 EEVVPXGMSY 10
    ||| | | |
    20 EEVFPPLANY 29

Db

RESULT 14
US-08-460-744-4
Sequence 4, Application US/08460744
Patent No. 6107541
GENERAL INFORMATION:
APPLICANT: Arnold, Andrew
TITLE OF INVENTION: Pradi Cyclin and its CDNA
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSER: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,744
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McConathy, Evelyn H.
REGISTRATION NUMBER: 35,279
REFERENCE/DOCKET NUMBER: 0609,4070005
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 152 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
```

```
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-460-744-4

Query Match
Best Local Similarity 61.5%; Score 32; DB 3; Length 152;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Cy 1 EEVVPXGMSY 10
    ||| | | |
    20 EEVFPPLANY 29

Db

RESULT 15
US-07-667-711B-4
Sequence 4, Application US/07667711B
Patent No. 6110700
GENERAL INFORMATION:
APPLICANT: ARNOLD, ANDREW
TITLE OF INVENTION: Pradi Cyclin and Its CDNA
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSER: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVE., NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/667,711B
FILING DATE: 11-MAR-1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MCPHAIL, DONALD R.
REGISTRATION NUMBER: 35,811
REFERENCE/DOCKET NUMBER: 0609,4070000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 152 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
US-07-667-711B-4

Query Match
Best Local Similarity 61.5%; Score 32; DB 3; Length 152;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Cy 1 EEVVPXGMSY 10
    ||| | | |
    20 EEVFPPLANY 29

Db

Search completed: December 22, 2003, 16:43:46
Job time : 10.1333 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 16:40:14 ; Search time 20.8667 Seconds
(without alignments)
98.451 Million cell updates/sec

US-09-909-164-47

Title: US-09-909-164-47
Perfect score: 52
Sequence: 1 BEVVPXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_NEW_PUB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	38	73.1	3472	12	US-10-029-120-4
2	38	73.1	3472	14	US-10-027-806-4
3	38	73.1	3472	14	US-10-034-623-4
4	38	73.1	3472	15	US-10-027-801-4
5	36	69.2	153	12	US-10-029-386-32486
6	35	67.3	236	12	US-10-029-386-32076
7	34	65.4	290	12	US-10-094-749-2565
8	34	65.4	947	15	US-10-101-464A-73
9	34	65.4	1163	12	US-10-122-067-4
10	34	65.4	1499	12	US-10-096-534-67
11	34	65.4	1499	12	US-10-122-067-2
12	33	63.5	99	15	US-10-106-598-7477
13	33	63.5	426	11	US-10-214-766-43
14	33	63.5	478	11	US-09-992-600A-108
15	33	63.5	478	11	US-09-924-340-108

16	33	63.5	478	11	US-09-746-783-184	Sequence 184, App
17	33	63.5 <td>478 <td>12</td> <td>US-09-992-0958-108</td> <td>Sequence 108, App</td> </td>	478 <td>12</td> <td>US-09-992-0958-108</td> <td>Sequence 108, App</td>	12	US-09-992-0958-108	Sequence 108, App
18	33	63.5 <td>478 <td>12</td> <td>US-10-154-678-108</td> <td>Sequence 108, App</td> </td>	478 <td>12</td> <td>US-10-154-678-108</td> <td>Sequence 108, App</td>	12	US-10-154-678-108	Sequence 108, App
19	33	63.5 <td>478 <td>12</td> <td>US-09-999-570-108</td> <td>Sequence 108, App</td> </td>	478 <td>12</td> <td>US-09-999-570-108</td> <td>Sequence 108, App</td>	12	US-09-999-570-108	Sequence 108, App
20	33	63.5 <td>478 <td>15</td> <td>US-10-000-489-108</td> <td>Sequence 108, App</td> </td>	478 <td>15</td> <td>US-10-000-489-108</td> <td>Sequence 108, App</td>	15	US-10-000-489-108	Sequence 108, App
21	33	63.5 <td>478 <td>15</td> <td>US-10-000-986-108</td> <td>Sequence 108, App</td> </td>	478 <td>15</td> <td>US-10-000-986-108</td> <td>Sequence 108, App</td>	15	US-10-000-986-108	Sequence 108, App
22	33	63.5 <td>478 <td>12</td> <td>US-10-094-749-3010</td> <td>Sequence 3010, App</td> </td>	478 <td>12</td> <td>US-10-094-749-3010</td> <td>Sequence 3010, App</td>	12	US-10-094-749-3010	Sequence 3010, App
23	33	63.5 <td>653</td> <td>11</td> <td>US-09-820-843A-26</td> <td>Sequence 26, App</td>	653	11	US-09-820-843A-26	Sequence 26, App
24	33	63.5 <td>1152</td> <td>16</td> <td>US-10-080-170-130</td> <td>Sequence 130, App</td>	1152	16	US-10-080-170-130	Sequence 130, App
25	33	63.5 <td>1394</td> <td>12</td> <td>US-10-116-275-261</td> <td>Sequence 261, App</td>	1394	12	US-10-116-275-261	Sequence 261, App
26	32	61.5 <td>10</td> <td>10</td> <td>US-09-947-387-66</td> <td>Sequence 66, App</td>	10	10	US-09-947-387-66	Sequence 66, App
27	32	61.5 <td>10</td> <td>12</td> <td>US-10-138-375-66</td> <td>Sequence 66, App</td>	10	12	US-10-138-375-66	Sequence 66, App
28	32	61.5 <td>36</td> <td>12</td> <td>US-10-199-820-241</td> <td>Sequence 241, App</td>	36	12	US-10-199-820-241	Sequence 241, App
29	32	61.5 <td>242</td> <td>12</td> <td>US-10-094-749-2076</td> <td>Sequence 2076, App</td>	242	12	US-10-094-749-2076	Sequence 2076, App
30	32	61.5 <td>254</td> <td>9</td> <td>US-09-778-927A-53</td> <td>Sequence 53, App</td>	254	9	US-09-778-927A-53	Sequence 53, App
31	32	61.5 <td>276</td> <td>12</td> <td>US-10-116-275-114</td> <td>Sequence 114, App</td>	276	12	US-10-116-275-114	Sequence 114, App
32	32	61.5 <td>289</td> <td>10</td> <td>US-09-919-497-54</td> <td>Sequence 54, App</td>	289	10	US-09-919-497-54	Sequence 54, App
33	32	61.5 <td>289</td> <td>14</td> <td>US-10-024-066-2</td> <td>Sequence 2, App</td>	289	14	US-10-024-066-2	Sequence 2, App
34	32	61.5 <td>289</td> <td>14</td> <td>US-10-024-066-4</td> <td>Sequence 4, App</td>	289	14	US-10-024-066-4	Sequence 4, App
35	32	61.5 <td>292</td> <td>12</td> <td>US-10-116-275-275</td> <td>Sequence 275, App</td>	292	12	US-10-116-275-275	Sequence 275, App
36	32	61.5 <td>295</td> <td>10</td> <td>US-09-925-300-1061</td> <td>Sequence 1061, App</td>	295	10	US-09-925-300-1061	Sequence 1061, App
37	32	61.5 <td>529</td> <td>9</td> <td>US-09-923-304-4</td> <td>Sequence 4, App</td>	529	9	US-09-923-304-4	Sequence 4, App
38	32	61.5 <td>529</td> <td>9</td> <td>US-09-925-731-2</td> <td>Sequence 2, App</td>	529	9	US-09-925-731-2	Sequence 2, App
39	32	61.5 <td>691</td> <td>15</td> <td>US-10-101-921-4</td> <td>Sequence 4, App</td>	691	15	US-10-101-921-4	Sequence 4, App
40	32	61.5 <td>1377</td> <td>9</td> <td>US-09-815-242-10384</td> <td>Sequence 10384, App</td>	1377	9	US-09-815-242-10384	Sequence 10384, App
41	32	61.5 <td>1377</td> <td>12</td> <td>US-10-287-274-467</td> <td>Sequence 467, App</td>	1377	12	US-10-287-274-467	Sequence 467, App
42	32	61.5 <td>2799</td> <td>14</td> <td>US-10-151-736-4</td> <td>Sequence 4, App</td>	2799	14	US-10-151-736-4	Sequence 4, App
43	31	59.6	53	10	US-09-764-847-878	Sequence 878, App
44	31	59.6	53	15	US-10-092-154-878	Sequence 878, App
45	31	59.6	59	10	US-09-948-080-14	Sequence 14, App

ALIGNMENTS

RESULT: 1
US-10-029-120-4
Sequence 4, Application US/10029120
Publication No. US20030175708A1
GENERAL INFORMATION:
APPLICANT: Swanson, Ronald V.
APPLICANT: Feldman, Robert A.
APPLICANT: Schleper, Christa
TITLE OR INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM
FILE REFERENCE: DCCP. 002A
CURRENT APPLICATION NUMBER: US/10/029,120
PRIOR FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408, 020
PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
NUMBER OF SEQ ID NOS: 123
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 4
LENGTH: 3472
TYPE: PRT
ORGANISM: Cenarchaeum symbiosum
US-10-029-120-4

Query Match 73.1%; Score 38; DB 12; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.6e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

DB 2294 BEVVPXGMSYS 11
2294 BEVVPXGMSYS 2304

RESULT: 2
US-10-027-806-4
Sequence 4, Application US/10027806
Publication No. US20020160476A1
GENERAL INFORMATION:
APPLICANT: Swanson, Ronald V.
APPLICANT: Feldman, Robert A.

```

; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOIP.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-806-4

Query Match      73.1%; Score 38; DB 14; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.6e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY      1 BEVVPXGMSYS 11
DB      2294 EDVIPRGISFS 2304

RESULT 3
US-10-034-623-4
; Sequence 4, Application US/10034623
; Publication No. US20020198365A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOIP.002A
; CURRENT APPLICATION NUMBER: US/10/034,623
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-034-623-4

Query Match      73.1%; Score 38; DB 14; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.6e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY      1 BEVVPXGMSYS 11
DB      2294 EDVIPRGISFS 2304

RESULT 4
US-10-027-801-4
; Sequence 4, Application US/10027801
; Publication No. US20030054364A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOIP.002A
; CURRENT APPLICATION NUMBER: US/10/027,801
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0

```

```

; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-801-4

Query Match      73.1%; Score 38; DB 15; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.6e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY      1 BEVVPXGMSYS 11
DB      2294 EDVIPRGISFS 2304

RESULT 5
US-10-029-386-32486
; Sequence 32486, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: ABOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32486
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP002994.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: SWISSPROT HIT: P43363, EVALU8 5.00e-46
US-10-029-386-32486

Query Match      69.2%; Score 36; DB 12; Length 153;
Best Local Similarity 70.0%; Pred. No. 12;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 BEVVPXGMSY 10
DB      15 KEVVPXGMSY 24

RESULT 6
US-10-029-386-32076
; Sequence 32076, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: ABOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32076
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

```

```
; OTHER INFORMATION: MAP TO 238050.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.3
; OTHER INFORMATION: SWISSPROT HIT: P15822, EVALU8 1.00e-125
US-10-029-386-32076

Query Match      67.3%; Score 35; DB 12; Length 236;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      3 VVPXGMSYS 11
      ||| ||| |||
      80 VVPAGLTYS 88

RESULT 7
US-10-094-749-2565
; Sequence 2565, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHUKA, ICHIRO
; APPLICANT: SEKI, NAOMIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2565
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2565

Query Match      65.4%; Score 34; DB 12; Length 290;
Best Local Similarity 66.7%; Pred. No. 66;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 BEVVPXGMS 9
      ||| ||| |||
      35 BEIVPMGIS 43

RESULT 8
US-10-101-464A-73
; Sequence 73, Application US/10101464A
; Publication No. US20030046728A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signaling
; FILE REFERENCE: 11000.1020c2
; CURRENT APPLICATION NUMBER: US/10/101,464A
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 73
; LENGTH: 947
; TYPE: PRT
; ORGANISM: Pinus radiata
US-10-101-464A-73

Query Match      65.4%; Score 34; DB 15; Length 947;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      3 VVPXGMSYS 11
      ||| ||| |||
      686 VMPGSGISYS 694

RESULT 9
US-10-122-067-4
; Sequence 4, Application US/10122067
; Publication No. US20030165883A1
; GENERAL INFORMATION:
; APPLICANT: Roy A.J. Curtiss
; TITLE OF INVENTION: 27091, A PHOSPHOLIPID TRANSPORTING
; TITLE OF INVENTION: ATPASE MOLECULE AND USRS THEREFOR
; FILE REFERENCE: MP101-058P1R
; CURRENT APPLICATION NUMBER: US/10/122,067
; CURRENT FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/283434
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1163
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-122-067-4

Query Match      65.4%; Score 34; DB 12; Length 1163;
Best Local Similarity 72.7%; Pred. No. 3.1e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 BEVVPXGMSYS 11
      ||| ||| |||
      133 BEVPPRGISYS 143

RESULT 10
US-10-096-534-67
; Sequence 67, Application US/10096534
; Publication No. US20030166887A1
; GENERAL INFORMATION:
; APPLICANT: The Brigham and Women's Hospital, Inc.
; APPLICANT: Yates, Karen
; APPLICANT: Mizuno, Shuichi
; APPLICANT: Glowacki, Julie
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF SKELETAL DEGENERATION CONDITIONS
```

```
FILE REFERENCE: B0801/7244/XA/ERP
CURRENT APPLICATION NUMBER: US/10/096,534
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: US 60/274,980
PRIOR FILING DATE: 2001-03-12
NUMBER OF SEQ ID NOS: 79
SOFTWARE: PatentIn version 3.0
SEQ ID NO 67
LENGTH: 1499
TYPE: PRT
ORGANISM: Homo sapiens
US-10-096-534-67

Query Match      65.4% Score 34; DB 12; Length 1499;
Best Local Similarity 72.7%; Pred. No. 4,2e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 469 BEVPRGGSVS 479

RESULT 11
US-10-122-067-2
Sequence 2, Application US/10122067
Publication No. US2003016583A1
GENERAL INFORMATION:
APPLICANT: ROY A.J. CURTIS
TITLE OF INVENTION: A PHOSPHOLIPID TRANSPORTING
TITLE OF INVENTION: ATPASE MOLECULE AND USES THEREFOR
FILE REFERENCE: MP101-058PIR
CURRENT APPLICATION NUMBER: US/10/122,067
CURRENT FILING DATE: 2002-04-12
PRIOR APPLICATION NUMBER: 60/283434
PRIOR FILING DATE: 2001-04-12
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 1499
TYPE: PRT
ORGANISM: Homo sapiens
US-10-122-067-2

Query Match      65.4% Score 34; DB 12; Length 1499;
Best Local Similarity 72.7%; Pred. No. 4,2e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 469 BEVPRGGSVS 479

RESULT 12
US-10-106-698-7477
Sequence 7477, Application US/10106698
Publication No. US20030109690A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
FILE REFERENCE: PA005PI
CURRENT APPLICATION NUMBER: US/10/106,698
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 7477
LENGTH: 99
TYPE: PRT
```

```
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (3)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (13)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (14)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (39)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (40)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (57)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (58)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (63)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (65)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (82)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (93)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-7477

Query Match      63.5% Score 33; DB 15; Length 99;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
DB 73 LVVPGISYS 81

RESULT 13
US-10-214-766-43
Sequence 43, Application US/10214766
Publication No. US20030084473A1
GENERAL INFORMATION:
APPLICANT: Gocal, Greg
TITLE OF INVENTION: NON-TRANSGENIC HERBICIDE RESISTANT PLANTS
FILE REFERENCE: CA1138
CURRENT APPLICATION NUMBER: US/10/214,766
CURRENT FILING DATE: 2002-08-09
PRIOR APPLICATION NUMBER: US 60/311,734
PRIOR FILING DATE: 2001-08-09
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.2
SEQ ID NO 43
LENGTH: 426
TYPE: PRT
ORGANISM: Vibrio cholerae
US-10-214-766-43

Query Match      63.5% Score 33; DB 15; Length 426;
Best Local Similarity 60.0%; Pred. No. 1,6e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
DB 223 EFVIPAGOSY 232
```

RESULT 14

US-09-992-600A-108
; Sequence 108, Application US/09992600A
; Publication No. US20030027161A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US4.DIV
; CURRENT APPLICATION NUMBER: US/09/992,600A
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: JPatent
; SEQ ID NO 108
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-992-600A-108

Query Match

63.5%; Score 33; DB 11; Length 478;
Best Local Similarity 60.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 2 EVVPGMSYS 11

Db 239 EVAPAGASYN 248

RESULT 15

US-09-924-340-108
; Sequence 108, Application US/09924340
; Publication No. US20030027248A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US2.RRG
; CURRENT APPLICATION NUMBER: US/09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 108
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-924-340-108

Query Match

63.5%; Score 33; DB 11; Length 478;
Best Local Similarity 60.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 2 EVVPGMSYS 11

Db 239 EVAPAGASYN 248

Search completed: December 22, 2003, 17:32:46
Job time : 21.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:24:36 ; Search time 9.06667 Seconds

(without alignments)
116.675 Million cell updates/sec

Title: US-09-909-164-47
Perfect score: 52
Sequence: 1 BEVVPXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 segs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	73.1	3472	2	hypothetical 367K
2	37	71.2	840	2	probable sulfate p
3	37	71.2	877	2	sulfate permease -
4	36	69.2	102	2	VI protein - tobac
5	36	69.2	1498	2	DNA segregation AT
6	35	67.3	225	2	hypothetical prote
7	35	67.3	425	2	hypothetical prote
8	35	67.3	670	2	zinc finger protei
9	35	67.3	749	2	topoisomerase IV s
10	35	67.3	2717	2	DNA-binding protei
11	34	65.4	156	2	hypothetical prote
12	34	65.4	252	2	cell division inh
13	34	65.4	544	2	probable ABC sub
14	33	63.5	94	2	hypothetical prote
15	33	63.5	116	2	50S ribosomal prot
16	33	63.5	165	2	hypothetical prote
17	33	63.5	253	2	hypothetical prote
18	33	63.5	259	2	hypothetical prote
19	33	63.5	284	2	hypothetical prote
20	33	63.5	298	2	beta-ketocyl-ACP
21	33	63.5	368	2	hypothetical prote
22	33	63.5	466	2	3-phosphoshikimate
23	33	63.5	466	2	cdcs37 protein - fi
24	33	63.5	653	2	iron(III) ABC tran
25	33	63.5	1028	2	bacteriocin BCN5 -
26	33	63.5	1080	2	ATP-dependent DNA
27	33	63.5	1152	2	conserved hypotet
28	33	63.5	1394	2	transforming growth
29	33	63.5	1401	2	DNA-directed RNA p

30	33	63.5	1548	2	T04456	hypothetical prote
31	32	61.5	1712	2	A82651	masking protein pr
32	32	61.5	84	2	B97333	hypothetical prote
33	32	61.5	175	2	PQ0616	transport protein
34	32	61.5	223	2	T01457	rho protein GDP-di
35	32	61.5	279	2	B72481	hypothetical prote
36	32	61.5	288	2	JC4011	hypothetical prote
37	32	61.5	288	2	I58372	cyclin D2 - rat
38	32	61.5	289	2	A41984	cyclin D2 - rat
39	32	61.5	289	2	A42822	cyclin D2 - mouse
40	32	61.5	291	2	S57922	cyclin D1 - human
41	32	61.5	291	2	S57925	cyclin D2 - Africa
42	32	61.5	291	2	JC4579	cyclin D2 - chicken
43	32	61.5	291	2	S62730	cyclin D1 - zebra
44	32	61.5	292	2	B42822	cyclin D3 - human
45	32	61.5	295	2	A38977	cyclin D1 - human

ALIGNMENTS

RESULT 1

T31308
hypothetical 367K protein - Cenarchaeum symbiosum
C/Species: Cenarchaeum symbiosum
C/Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C/Accession: T31308
R/Schleper, C.; DeLong, E.F.; Preston, C.M.; Feldman, R.A.; Wu, K.Y.; Swanson, R.V.
J. Bacteriol. 180, 5003-5009, 1998
A/Title: Genomic analysis reveals chromosomal variation in natural populations of the
A/Reference number: Z20994, MUID:98422450, PMID:978430
A/Accession: T31308
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-3472 <SCH>
A/Cross-references: EMBL:AF083072; NID:g3599393; PID:g3599394; PIDN:AA62699.1
C/Superfamily: Cenarchaeum symbiosum hypothetical 367K protein

Query Match 73.1% Score 38; DB 2; Length 3472;
Best Local Similarity 54.5%; Pred. No. 59;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 2294 EDVIPRGISFS 2304

RESULT 2

T39116
probable sulfate permease - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C/Accession: T39116
R/Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, November 1999
A/Reference number: Z21829
A/Accession: T39116
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-840 <HUN>
A/Cross-references: EMBL:ALJ32779; PIDN:CA860015.1; GSPDB:GN00066; SPDB:SPAC869.05C
C/Genetic: A/Experimental source: strain 972h-; cosmid c869
A/Map position: 1
A/Map position: 1

Query Match 71.2% Score 37; DB 2; Length 840;
Best Local Similarity 77.8%; Pred. No. 21;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 BEVVPXGMSYS 11
DB 135 VFPQGMSTYA 143

```

RESULT 3
T40413
Enolate permease - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40413
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Jimenez Martinez, J.
submitted to the EMBL Data Library, August 1998
A:Reference number: Z21926
A:Accession: T40413
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-877 <LVN>
A:Cross-references: EMBL:AL031261; PIDN:CAA20298.1; GSPDB:GN00067; SPDB:SPBCH7.02
A:Experimental source: strain 972h-, cosmid c3H7
A:Genetics:
A:Gene: SPDB:SPBCH7.02
A:Map position: 2

Query Match      71.2%  Score 37;  DB 2;  Length 877;
Best Local Similarity 77.8%  Pred. No. 22;
Matches 7;  Conservative 1;  Mismatches 1;  Indels 0;  Gaps 0;

QY      3  VVPXGMSYS 11
      ||| |||
      148 VVPXGMSYA 156

RESULT 4
A42452
V1 protease - tobacco yellow dwarf virus (strain Australia)
C:Species: tobacco yellow dwarf virus
C:Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C:Accession: A42452
R:Morrill, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
Virology 187, 633-642, 1992
A:Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow
A:Reference number: A42452; MUID:92188538; PMID:1546458
A:Accession: A42452
A:Molecule type: DNA
A:Residues: 1-102 <MOR>
A:Cross-references: GB:M81103; NID:G335283; PIDN:AAA47947.1; PID:G335284

Query Match      69.2%  Score 36;  DB 2;  Length 102;
Best Local Similarity 60.0%  Pred. No. 3.5;
Matches 6;  Conservative 3;  Mismatches 1;  Indels 0;  Gaps 0;

QY      2  EVVPXGMSYS 11
      ||| |||
      7  QVVPXGMSYS 16

RESULT 5
B97355
DNA segregation Appase, Pter/SpoIIIE family, YUKA B. subtilis ortholog [imported] - Clo
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: B97355
R:Noelling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
A:Reference number: A86900; MUID:2135925; PMID:21359325
A:Accession: B97355
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1498 <KAY>
A:Cross-references: GB:AE001437; PIDN:AAK81629.1; PID:G15026814; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
A:Genetics:
A:Gene: CAC3709

```

```

Query Match      69.2%  Score 36;  DB 2;  Length 1498;
Best Local Similarity 60.0%  Pred. No. 63;
Matches 6;  Conservative 2;  Mismatches 2;  Indels 0;  Gaps 0;

QY      1  BEVVPXGMSY 10
      ||| |||
      1276 BQIVPGGLQY 1285

RESULT 6
S57810
hypothetical protein precursor (clone TPP1) - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
C:Accession: S57810
R:Milligan, S.B.; Gaaser, C.S.
Plant Mol. Biol. 28, 691-711, 1995
A:Title: Nature and regulation of ptc1-expressed genes in tomato.
A:Reference number: S57808; MUID:95375233; PMID:7647301
A:Accession: S57810
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-225 <MTL>
A:Cross-references: EMBL:U20592; NID:G924625; PIDN:AAA80497.1; PID:G924626
C:Superfamily: plant Kunitz-type proteinase inhibitor

Query Match      67.3%  Score 35;  DB 2;  Length 225;
Best Local Similarity 54.5%  Pred. No. 13;
Matches 6;  Conservative 3;  Mismatches 2;  Indels 0;  Gaps 0;

QY      1  BEVVPXGMSYS 11
      ||| |||
      32 DEVPNGKTYA 42

RESULT 7
T24111
hypothetical protein R10D12.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24111
R:Percy, C.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19842
A:Accession: T24111
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-425 <WTL>
A:Cross-references: EMBL:Z81109; PIDN:CAB03241.1; GSPDB:GN00023; CESP:R10D12.10
A:Experimental source: clone R10D12
A:Genetics:
A:Gene: CESP:R10D12.10
A:Map position: 5
A:introns: 23/3; 56/3; 113/3; 257/2

Query Match      67.3%  Score 35;  DB 2;  Length 425;
Best Local Similarity 50.0%  Pred. No. 26;
Matches 5;  Conservative 3;  Mismatches 2;  Indels 0;  Gaps 0;

QY      1  BEVVPXGMSY 10
      ||| |||
      335 BQIVPGGLQY 344

RESULT 8
S22293
zinc finger protein AT-BP2 - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 20-Sep-1999
C:Accession: S22293; I78656
R:Mitchelemore, C.; Traboni, C.; Cortese, R.
Nucleic Acids Res. 19, 141-147, 1991

```

A/Title: Isolation of two cDNAs encoding zinc finger proteins which bind to the alpha 1-
 A/Reference number: 158280; MUID:91187610; PMID:1901405
 A/Accession: S22293
 A/Status: nucleic acid sequence not shown
 A/Molecule type: mRNA
 A/Residues: 1-670 <MT>
 A/Cross-references: EMBL:X54250; NID:957519; PIDN:CAA3151.1; PID:957520
 A/Note: the authors did not translate the codon for residue 1
 C/Superfamily: HIV-SP2 enhancer-binding protein
 C/Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 670;
 Best Local Similarity 66.7%; Pred. No. 43;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 EVVPXGMSY 11
 ||| |||
 Db 376 VVPAGLTY 384

RESULT 9

H82691
 C/Species: Xylella fastidiosa
 C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C/Accession: H82691
 R/Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen-
 Nature 406, 151-157, 2000
 A/Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A/Reference number: A82515; MUID:20365717; PMID:10910347
 A/Note: for a complete list of authors see reference number A59328 below
 A/Accession: H82691
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-749 <STM>
 A/Cross-references: GB:AB003967; GB:AE003849; NID:9106347; PIDN:PAF84162.1; GSPDB:GN001
 A/Experimental source: strain 9a5c
 R/Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvares, R.; A-
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre, H-
 ab-Neto, E.; Docena, C.; El-Dorty, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Freaga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
 J.D.; Junqueira, M.L.; Kemp, E.L.; Kitajima, J.P.; Krieger, J.B.; Kuramae, E.B.; Laigt
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, R
 A.; Authors: Martins, E.M.F.; Matsumura, A.Y.; Menck, C.F.M.; Miracca, B.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
 Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A.; Authors: da Silva, A.C.R.; da Silva, P.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Tenhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A/Reference number: A59328
 A/Contents: annotation
 C/Genetics:
 A/Gene: XRF153
 C/Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain A/ phage T4 DNA topoisomerase (

Query Match 67.3%; Score 35; DB 2; Length 749;
 Best Local Similarity 77.8%; Pred. No. 48;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
 ||| |||
 Db 526 EVDPSGMSY 534

RESULT 10

A34203
 DNA-binding protein PRDII-BP1 - human
 N/Alternate names: major histocompatibility complex enhancer-binding protein 1
 C/Species: Homo sapiens (man)
 C/Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 20-Sep-1999
 C/Accession: A34203; A34779
 R/Pan, C.M.; Maniatis, T.
 Genes Dev. 4, 29-42, 1990

A/Title: A DNA-binding protein containing two widely separated zinc finger motifs that
 A/Reference number: A34203; MUID:90169514; PMID:2106471
 A/Accession: A34203
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-2717 <PAN>
 A/Cross-references: EMBL:X51435; NID:938017; PIDN:CAA5798.1; PID:938018
 R/Baldwin Jr., A.S.; Leclaire, K.P.; Singh, H.; Sharp, P.A.
 Mol. Cell. Biol. 10, 1406-1414, 1990
 A/Title: A large protein containing zinc finger domains binds to related sequence elem
 A/Reference number: A34779; MUID:90205817; PMID:2108316
 A/Accession: A34779
 A/Status: preliminary; nucleic acid sequence not shown
 A/Molecule type: mRNA
 A/Residues: 801-1072, 'N', 1074-1166, 'K', 1170-1225, 'V', 1227-1434, 'N', 1436-1607, 'I', 1609-
 A/Cross-references: GB:M32019
 C/Superfamily: HIV-SP2 enhancer-binding protein
 C/Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 2717;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 EVVPXGMSY 11
 ||| |||
 Db 2405 VVPAGLTY 2413

RESULT 11

H54619
 A/Title: A DNA-binding protein YOR013w - yeast (Saccharomyces cerevisiae)
 N/Alternate names: hypothetical protein O2612; hypothetical protein YOL303.3
 C/Species: Saccharomyces cerevisiae
 C/Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
 C/Accession: S54619; S66879
 R/de Haan, M.; Maarse, A.C.; Grivell, L.A.
 submitted to the EMBL Data Library, May 1995
 A/Reference number: S54617
 A/Accession: S54619
 A/Molecule type: DNA
 A/Residues: 1-156 <DEH>
 A/Cross-references: EMBL:X87331; NID:91041652; PIDN:CAA60762.1; PID:9829123
 R/de Haan, M.; Grivell, L.A.; Maarse, A.C.
 submitted to the Protein Sequence Database, July 1996
 A/Reference number: S66877
 A/Accession: S66879
 A/Molecule type: DNA
 A/Residues: 1-156 <DEW>
 A/Cross-references: EMBL:Z74920; NID:91420109; PIDN:CAA99201.1; PID:91420111; MIPS:YOF
 A/Experimental source: strain S288C
 C/Genetics:
 A/Map position: 15R
 A/Cross-references: SGD:S0005539
 A/Map position: 15R
 C/Superfamily: hypothetical protein YOR013w

Query Match 65.4%; Score 34; DB 2; Length 156;
 Best Local Similarity 66.7%; Pred. No. 14;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
 ||| |||
 Db 50 EVMPDGMGY 58

RESULT 12

H69491
 cell division inhibitor (mind-2) homolog - Archaeoglobus fulgidus
 C/Species: Archaeoglobus fulgidus
 C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 26-Aug-1999
 C/Accession: H69491
 R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dods
 R.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirtress, B.
 Glodex, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997
A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: H69491
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-252 <KLE>
A:Cross-references: GB:AB000970; GB:AE000782; NID:G2689293; PIDN:AB89318.1; PID:G264860
C:Superfamily: cell division inhibitor mind

Query Match 65.4%; Score 34; DB 2; Length 252;
Best Local Similarity 75.0%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 EVVFXGMS 9
DB 81 EVVFXGMS 88

RESULT 13

C82900
Probable ABC substrate-binding protein, iron UUS9 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: C82900
R:Glass, J.I.; Leikowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Caswell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mit
A:Reference number: A82870
A:Accession: C82900
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-544 <GRL>
A:Cross-references: GB:AE002133; GB:AP222894; NID:G6899339; PIDN:AA730768.1; GSPDB:GN001
A:Experimental source: Beroovar 3; Diovat 1
C:Genetics:
A:Gene: ABCsbp-5; UUS9
A:Genetic code: SGC3

Query Match 65.4%; Score 34; DB 2; Length 544;
Best Local Similarity 70.0%; Pred. No. 55;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EEVFXGMSY 10
DB 135 EEVFXGMSY 144

RESULT 14

I40758
hypothetical protein 1 - Campylobacter jejuni (fragment)
C:Species: Campylobacter jejuni
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
C:Accession: I40758; S47317
R:Hani, B.K.; Chan, V.L.
J. Bacteriol. 177, 2396-2402, 1995
A:Title: Expression and characterization of Campylobacter jejuni benzoyl-glycine amidohy
A:Reference number: I40758; MUID:95247673; PMID:7730270
A:Accession: I40758
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-94 <RES>
A:Cross-references: EMBL:Z36940; NID:G535805; PIDN:CA85392.1; PID:G535806

Query Match 63.5%; Score 33; DB 2; Length 94;
Best Local Similarity 55.6%; Pred. No. 14;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 EVVFXGMSY 10
DB 26 EVVFXGMSY 34

RESULT 15

B90544
50S ribosomal protein L20 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: B90544
R:Chamand, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pu
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: B90544
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-116 <KUR>
A:Cross-references: GB:AL445566; PID:G14089674; PIDN:CAC13434.1; GSPDB:GN00153
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MPRU_2610
A:Genetic code: SGC3
C:Superfamily: Escherichia coli ribosomal protein L20

Query Match 63.5%; Score 33; DB 2; Length 116;
Best Local Similarity 77.8%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 VVFXGMSYS 11
DB 68 VVFXGMSYS 76

Search completed: December 22, 2003, 17:45:02
Job time : 10.0667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:43:51 / Search time 4.6 Seconds

(without alignments)
112.455 Million cell updates/sec

Title: US-09-909-164-47
Perfect score: 52
Sequence: 1 BEVVPKGMYSYS 11

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	73.1	1058	1 CARB_FUSNN	Q8R86 fusobacteri
2	37	71.2	877	1 SUIH_SCHPO	O74377 schizosacch
3	36	69.2	102	1 Y1IK_TYDVA	P31619 tobacco yel
4	36	69.2	1498	1 Y1A9_CIOAB	O04351 clostridium
5	35	67.3	2717	1 ZEP1_HUMAN	P15822 homo sapien
6	34	65.4	788	1 CY14_NEUCR	P33622 neurospora
7	34	65.4	1499	1 A10C_HUMAN	O60332 homo sapien
8	33	63.5	116	1 RLZ0_MYCPU	O98QV0 mycoplasma
9	33	63.5	165	1 YJ49_ARCFU	O28330 archaeoglob
10	33	63.5	253	1 Y990_CAMJE	P45488 campylobact
11	33	63.5	280	1 CTX3_MOUSE	Q34387 mus musculu
12	33	63.5	426	1 AROA_VIBCH	Q9K7B0 vibrio chol
13	33	63.5	466	1 CC37_SCHPO	O94740 schizosacch
14	33	63.5	478	1 GSR2_HUMAN	Q9NZM5 homo sapien
15	33	63.5	890	1 BCNS_CLOPE	P08696 clostridium
16	33	63.5	1389	1 LTR8_MOUSE	O8CG18 mus musculu
17	33	63.5	1394	1 LTR8_HUMAN	P22064 homo sapien
18	33	63.5	1401	1 RPOC_VIBCH	Q9K7B2 vibrio chol
19	33	63.5	1595	1 LTR1_HUMAN	Q14766 homo sapien
20	33	63.5	1712	1 LTR1_RAT	O00918 rattus norv
21	33	63.5	1713	1 LTR1_MOUSE	O8CG19 mus musculu
22	32	61.5	288	1 CGD2_RAT	O04827 rattus norv
23	32	61.5	289	1 CGD2_HUMAN	P30279 homo sapien
24	32	61.5	289	1 CGD2_MOUSE	P30280 mus musculu
25	32	61.5	291	1 CGD1_BRABE	Q90459 brachydanio
26	32	61.5	291	1 CGD1_XENILA	P50755 xenopus lae
27	32	61.5	291	1 CGD2_CHICK	P49706 gallus gall
28	32	61.5	291	1 CGD2_XENILA	P53782 xenopus lae
29	32	61.5	292	1 CGD1_CHICK	P55169 gallus gall
30	32	61.5	292	1 CGD3_HUMAN	P30281 homo sapien
31	32	61.5	295	1 CGD1_HUMAN	P24388 homo sapien
32	32	61.5	295	1 CGD1_MOUSE	P25322 mus musculu
33	32	61.5	295	1 CGD1_RAT	P39948 rattus norv

34	32	61.5	427	1 TOLB_HAEIN	P44677 haemophilus
35	32	61.5	529	1 ENP3_HUMAN	O75355 homo sapien
36	32	61.5	691	1 S216_HUMAN	O9Y616 homo sapien
37	32	61.5	719	1 GSP_CRITA	P90518 crictridia f
38	32	61.5	726	1 PRTF_HSV6U	P52384 human herpe
39	32	61.5	759	1 SCTL_YEAST	P32784 saccharomyc
40	32	61.5	920	1 EDD_RAT	O62671 rattus norv
41	32	61.5	993	1 VIA_TAV	P28931 tomato aspe
42	32	61.5	1377	1 RHA5_ECOLI	P16916 escherichia
43	32	61.5	1397	1 RHSC_ECOLI	P16918 escherichia
44	32	61.5	1411	1 RHSE_ECOLI	P16917 escherichia
45	32	61.5	2799	1 EDD_HUMAN	O95071 homo sapien

ALIGNMENTS

```

RESULT 1
ID CARB_FUSNN STANDARD; PRT; 1058 AA.
AC Q8R86;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) [Carbamoyl]-
DE phosphate synthetase ammonia chain).
GN CARB OR FN0442.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SOURCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; Pubmed=11899109;
RA Kapralov V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman B., Bernal A.,
RA Larsen N., D'Souza M., Malinas T., Pusch G., Haselkorn R.,
RA Fomstein M., Kyrides N., Overbeek R.;
RT genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.
J. Bacteriol. 184:2005-2018(2002).
-1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
phosphate + L-glutamate + carbamoyl phosphate.
-1- COFACTOR: Binds 3 manganese ions per subunit. (By similarity).
-1- PATHWAY: Arginine biosynthesis.
-1- PATHWAY: Pyrimidine biosynthesis; first step.
-1- SUBUNIT: Composed of two chains; the small (or glutamine) chain
promotes the hydrolysis of glutamine to ammonia, which is used by
the large (or ammonia) chain to synthesize carbamoyl phosphate (by
similarity).
-1- SIMILARITY: BELONGS TO THE CARB FAMILY.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL, AB010554, AA094625.1, ALT_INTT.
CC HMAAP; MF_01210; - 1.
DR InterPro: IPR006275; CARB_L_glu.
DR InterPro: IPR005483; CPase_L.
DR InterPro: IPR005479; CPase_L_D2.
DR InterPro: IPR005480; CPase_L_D3.
DR InterPro: IPR005481; CPase_L_N.
DR InterPro: IPR004362; MGS_Like.
DR Pfam; PF02089; CPase_L_Chain; 2.
DR Pfam; PF02786; CPase_L_D2; 2.
DR Pfam; PF02787; CPase_L_D3; 1.

```

DR Pfam; PF02142; MGS; 1.
 DR PRINTS; PR00098; CPSASE.
 DR TIGRFAMs; TIGR01369; CPSaseII_1rg; 1.
 DR PROSITE; PS00866; CPSASE 1; 2.
 DR PROSITE; PS00867; CPSASE 2; 2.
 KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 KM ATP-binding; Manganese; Complete proteome.
 FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
 FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
 FT REPEAT 1 546
 FT REPEAT 547 1058
 FT NP_BIND 153 210 ATP (POTENTIAL).
 FT NP_BIND 302 352 ATP (POTENTIAL).
 FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
 FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
 FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
 FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
 FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
 SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1B39F CRC64;
 Query Match 73.1%; Score 38; DB 1; Length 1058;
 Best Local Similarity 60.0%; Pred. No. 6.3;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPGMSYS 11
 Db 190 EIVNGMYS 199

RESULT 2
 ID SUHL SCHPO STANDARD; PRT; 877 AA.
 AC 074377;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable sulfate permease C3H7.02.
 OS SPEC3H7.02.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=977;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Holtroyd S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Moore P., Moule S., Mungall K., Murphy L., Mablet D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Voiclaert G., Aert R., Robben G., Grymoprez B.,
 RA Weidens J., Vancraele E., Rieger M., Scheffer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Leizer E., Moesli D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Weller H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleau V., Mottier S.,
 RA Galibert F., Aves S.-J., Xiang Z., Hunt C., Moore K., Huert S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
 RA Dues R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;

RT "The genome sequence of Schizosaccharomyces pombe."
 RL Nature 415:871-880(2002).
 CC -1- FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE INTO THE CELL (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE SLC26A FAMILY OF TRANSPORTERS.
 CC -1- SIMILARITY: Contains 1 STAS domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.embnet.org/announcements> or send an email to license@ebi-embnet.org).
 CC -----
 DR EMBL; AL031261; CAA20298.1; -.
 DR PIR; T40413; T40413.
 DR GeneDB; Sprobe; SPBC3H7.02; -.
 DR InterPro; IPR002645; STAS.
 DR InterPro; IPR001902; Sulph_transpt.
 DR Pfam; PF01740; STAS; 1.
 DR Pfam; PF00916; Sulfate_transp; 1.
 DR TIGRFAMs; TIGR00815; sulp; 1.
 DR PROSITE; PS01130; SLC26A; 1.
 DR PROSITE; PS50801; STAS; 1.
 KM Transport; Transmembrane.
 FT TRANSMEM 133 153 POTENTIAL.
 FT TRANSMEM 161 181 POTENTIAL.
 FT TRANSMEM 186 206 POTENTIAL.
 FT TRANSMEM 221 241 POTENTIAL.
 FT TRANSMEM 243 263 POTENTIAL.
 FT TRANSMEM 292 312 POTENTIAL.
 FT TRANSMEM 329 349 POTENTIAL.
 FT TRANSMEM 384 404 POTENTIAL.
 FT TRANSMEM 424 444 POTENTIAL.
 FT TRANSMEM 461 481 POTENTIAL.
 FT TRANSMEM 484 504 POTENTIAL.
 FT TRANSMEM 518 538 POTENTIAL.
 FT TRANSMEM 543 563 POTENTIAL.
 FT DOMAIN 594 747 STAS.
 SQ SEQUENCE 877 AA; 96373 MW; 56959A0493371B43 CRC64;
 Query Match 71.2%; Score 37; DB 1; Length 877;
 Best Local Similarity 77.8%; Pred. No. 8.4;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPGMSYS 11
 Db 148 VVPGMSYS 156

RESULT 3
 ID YLIK TYDVA STANDARD; PRT; 102 AA.
 AC P31619;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-1993 (Rel. 27, Last annotation update)
 DE Hypothetical 11.2 kDa protein.
 OS Hypothetical 11.2 kDa protein.
 OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
 OC Viruses; ssDNA viruses; Gemintoviridae; Maestrevirus.
 OX NCBI_TaxID=31599;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=92188536; PubMed=1546458;
 RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.B.;
 RT "The nucleotide sequence of the infectious cloned DNA component of
 RT tobacco yellow dwarf virus reveals features of gemintoviruses
 RT infecting monocotyledonous plants."
 RT Virology 187:633-642(1992).
 RL -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; M81103; AAA47947.1; -
DR PIR; A42452; A42452.
DR InterPro; IPR002621; Gemini_mov.
DR Pfam; PF01708; Gemini_mov; 1.
DR Hypothetical protein.
SQ SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF55B67 CRC64;

Query Match 69.2%; Score 36; DB 1; Length 102;
Best Local Similarity 60.0%; Pred. No. 1.4;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPGKMSYS 11
: ||| ||| :
Db 7 QVPSGGINYS 16

RESULT 4
Y1A9_CLOAB STANDARD; PRT; 1498 AA.
ID Y1A9_CLOAB
AC Q04351.

DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein CAC3709.
GN CAC3709.

OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
CC Clostridium.
NCBI_TaxID=1488;

[1]
SEQUENCE FROM N.A.

RP STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Neill J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucet-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838 (2001).

[2]
SEQUENCE OF 1-108 FROM N.A.

RP STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=93273706; PubMed=8501044;
RA Sauer U., Duerre P.;
RT "Sequence and molecular characterization of a DNA region encoding a
RT small heat shock protein of Clostridium acetobutylicum.";
RL J. Bacteriol. 175:3394-3400 (1993).
CC -1- SIMILARITY: Contains 2 FtsK/SpoIIIE domains.
CC -1- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
CC in positions 76 and 106.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; AE007866; AAK81629.1; -
DR EMBL; X65276; CA446379.1; ALT_FRAME.
DR PIR; B97355; B97355.
DR InterPro; IPR002543; FtsK_SpoIIIE.
DR Pfam; PF01580; FtsK_SpoIIIE; 2.

DR PROSITE; PS50901; FTSK_SPOIIIE; 2.
RX Hypothetical protein; ATP-binding; Complete proteome; Repeat.
FT DOMAIN 655 857 FTSK_SPOIIIE 1.
FT NP BIND 675 682 ATP (POTENTIAL).
FT DOMAIN 1001 1188 FTSK_SPOIIIE 2.
SQ SEQUENCE 1498 AA; 168968 MW; FFA2037A335A9649 CRC64;

Query Match 69.2%; Score 36; DB 1; Length 1498;
Best Local Similarity 60.0%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVPGKMSY 10
: ||| ||| :
Db 1276 EQKIPMGMSY 1285

RESULT 5
ZBP1_HUMAN STANDARD; PRT; 2717 AA.
ID ZBP1_HUMAN
AC P15832.

DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 40 (Human immunodeficiency virus type 1 enhancer-
DE binding protein 1) (HIV-BP1) (Major histocompatibility complex binding
DE protein 1) (MBP-1) (Positive regulatory domain II binding factor 1)
DE (PRDI1-BP1).

GN HIVBP1 OR ZNF40.
OS Homo sapiens (Human).
OC Bacteria; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;

[1]
SEQUENCE FROM N.A.
RX MEDLINE=90169514; PubMed=2106471;
RA Fan C.M., Maniatis T.;
RT "A DNA-binding protein containing two widely separated zinc finger
RT motifs that recognize the same DNA sequence.";
RL Genes Dev. 4:29-42 (1990).

[2]
STRUCTURE BY NMR OF 2113-2142.
RX MEDLINE=9106433; PubMed=2248949;
RA Omelchenko J.G., Clore G.M., Appella E., Sakaguchi K.,
RA Gronenborn A.M.;
RT "High-resolution three-dimensional structure of a single zinc finger
RT from a human enhancer binding protein in solution.";
RL Biochemistry 29:9324-9334 (1990).

[3]
STRUCTURE BY NMR OF 2087-2142.
RX MEDLINE=9232684; PubMed=1567844;
RA Omelchenko J.G., Clore G.M., Robien M., Sakaguchi K., Appella E.,
RA Gronenborn A.M.;
RT "High-resolution solution structure of the double Cys2His2 zinc
RT finger from the human enhancer binding protein MBP-1.";
RL Biochemistry 31:3907-3917 (1992).

-1- FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE
5'-GGGACTTTC-3' WHICH IS FOUND IN THE ENHANCER ELEMENTS OF
NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1.
IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS
OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I
MHC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT
IN T-CELL ACTIVATION.

-1- SUBCELLULAR LOCATION: Nuclear.
-1- INDUCTION: BY MITOGEN AND PHORBOL ESTER.
-1- DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY
SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH
ZINC-FINGER IN-BETWEEN.

-1- SIMILARITY: STRONG, TO HIVBP2.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).

 DR EMBL: X51435; CAA35798.1; -
 DR PDB: A34203; A34203.
 DR PDB: 3ZNF; 15-JAN-92.
 DR PDB: 4ZNF; 15-JAN-92.
 DR PDB: 1BBO; 31-OCT-93.
 DR TRANSFAC: T00497; -
 DR Genew: HGNC:4920; HIVEP1.
 DR MIM: 194540; -
 DR GO: GO:0005634; C:nucleus; TMS.
 DR GO: GO:0003677; F:DNA binding activity; TMS.
 DR InterPro: IPR007087; Znf_C2H2.
 DR Pfam: PF00096; zf-C2H2; 5.
 DR SMART: SM00355; ZNF_C2H2; 4.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 4.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 4.
 DR Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
 KW Nucleic protein; Repeat; 3D-structure.
 FT ZN_FING 406 428 C2H2-TYPE.
 FT ZN_FING 434 456 C2H2-TYPE.
 FT ZN_FING 958 981 C2H2-TYPE (POTENTIAL).
 FT ZN_FING 2087 2109 C2H2-TYPE.
 FT ZN_FING 2115 2139 C2H2-TYPE.
 FT DOMAIN 803 806 POLY-SER.
 FT STRAND 2088 2088
 FT TURN 2090 2092
 FT STRAND 2095 2095
 FT HELIX 2099 2108
 FT TURN 2109 2109
 FT STRAND 2115 2116
 FT STRAND 2123 2124
 FT HELIX 2127 2135
 SQ SEQUENCE 2717 AA; 297217 MW; D45D3CA951FEA561 CRC64;

Query Match 67.3%; Score 35; DB 1; Length 2717;
 Best Local Similarity 66.7%; Pred. No. 73;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVFXGMSYS 11
 DB 2405 VVPAGLTY 2413

RESULT 6
 CY14_NEUCR STANDARD; PRT; 788 AA.
 AC P23622;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cys14. permenase II.
 GN CYS-14.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91129256; PubMed=1825178;
 RA Kletter U.S., Jatal G., Fu Y.-H., Marzluf G.A.;
 RT "Nucleotide sequence, messenger RNA stability, and DNA recognition
 RT elements of cys-14, the structural gene for sulfate permenase II in
 RT Neurospora crassa";
 RL Biochemistry 30:1780-1787(1991).
 RN [2]
 RP PROBABLE REVISIONS.
 RX MEDLINE=94188926; PubMed=8140616;
 RA Sandal N.N., Marcher K.A.;
 RT "Similarities between a soybean nodulin, Neurospora crassa sulphate
 RT permenase II and a putative human tumour suppressor.";

RL Trends Biochem. Sci. 19:19-19(1994).
 CC -1- FUNCTION: UPTAKE OF SULFATE INTO THE CELL.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- INDICATION: Highly expressed, but only in cells subject to sulfur
 CC limitation, and it is turned on by the positive-acting Cys-3
 CC sulfur regulatory protein.
 CC -1- MISCELLANEOUS: SULFATE PERMEASE II IS MAINLY FOUND IN MYCELLA.
 CC -1- SIMILARITY: BELONGS TO THE SLC26A FAMILY OF TRANSPORTERS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).

 DR EMBL: M59167; AAA33615.1; ALT_SEQ.
 DR InterPro: IPR001902; Sulph_transp.
 DR Pfam: PF00916; Sulfate_transp; 1.
 DR TIGRPFAM: TIGR00815; sulp; 1.
 DR PROSITE: PS01130; SLC26A; 1.
 DR TransPort; Transmembrane; Glycoprotein.
 FT TRANSMEM 71 91 POTENTIAL.
 FT TRANSMEM 103 123 POTENTIAL.
 FT TRANSMEM 128 148 POTENTIAL.
 FT TRANSMEM 171 191 POTENTIAL.
 FT TRANSMEM 193 213 POTENTIAL.
 FT TRANSMEM 271 291 POTENTIAL.
 FT TRANSMEM 326 346 POTENTIAL.
 FT TRANSMEM 363 383 POTENTIAL.
 FT TRANSMEM 451 471 POTENTIAL.
 FT TRANSMEM 474 494 POTENTIAL.
 FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 788 AA; 87864 MW; 4FC604B60798C877 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 788;
 Best Local Similarity 66.7%; Pred. No. 34;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVFXGMSYS 11
 DB 90 VVFGMATA 98

RESULT 7
 ALOC_HUMAN STANDARD; PRT; 1499 AA.
 AC O60312; Q96914;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Potential phospholipid-transporting ATPase VC (BC 3.6.3.1) (ATPVC)
 DE (aminophospholipid translocase VC).
 GN Atp10C OR ATPVC OR KIA00566.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2125279; PubMed=11326269;
 RA Meguro M., Kaishiwagi A., Mitsuya K., Nakao M., Kondo I., Satoh S.,
 RA Ohshima M.;
 RT "A novel maternally expressed gene, Atp10C, encodes a putative
 RT aminophospholipid translocase associated with Angelman syndrome.";
 RL Nat. Genet. 28:19-20(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21313119; PubMed=11353404;
 RA Herzog L.B.K., Kim S.-J., Cook B.H., Jr., Ledbetter D.H.;
 RT "The human aminophospholipid-transporting ATPase gene Atp10C maps

RT adjacent to UBR3A and exhibits similar imprinted expression.";
 RL Am. J. Hum. Genet. 68:1501-1505(2001).
 RP [3]
 RC SEQUENCE OF 337-1499 FROM N.A.
 RX TISSUE=Brain;
 RA MEDLINE=96290545; PubMed=9628581;
 RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
 RA Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. IX.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 5:31-39(1998).
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -1- TISSUE SPECIFICITY: Wide expression, with highest levels in
 CC kidney, followed by lung, brain, prostate, testis, ovary, and
 CC small intestine.
 CC -1- DISEASE: Defects in ATP10C are a cause of Angelman syndrome (AS)
 CC [MIM:105830]; also known as 'happy puppet syndrome'. It is
 CC characterized by features of severe motor and intellectual
 CC retardation, microcephaly, ataxia, frequent jerky limb movements
 CC and flapping of the arms and hands, hypotonia, hyperactivity,
 CC hypopigmentation, seizures, absence of speech, frequent smiling
 CC and episodes of paroxysmal laughter, and an unusual facies
 CC characterized by macrostomia, a large mandible and open-mouthed
 CC expression, a great propensity for protruding the tongue ('tongue
 CC thrusting'), and an occipital groove.
 CC -1- SIMILARITY: Belongs to the cation transport ATPases family (P-type
 CC ATPases). Subfamily IV.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

RM Multigene family.
 FT DOMAIN 1 86
 FT TRANSFEM 87 106
 FT DOMAIN 107 110
 FT TRANSFEM 111 128
 FT DOMAIN 129 309
 FT TRANSFEM 310 332
 FT DOMAIN 333 362
 FT TRANSFEM 363 384
 FT DOMAIN 385 1087
 FT TRANSFEM 1088 1108
 FT DOMAIN 1109 1119
 FT TRANSFEM 1120 1140
 FT DOMAIN 1141 1170
 FT TRANSFEM 1171 1192
 FT DOMAIN 1193 1199
 FT TRANSFEM 1200 1222
 FT DOMAIN 1223 1228
 FT TRANSFEM 1229 1249
 FT DOMAIN 1250 1267
 FT TRANSFEM 1268 1292
 FT DOMAIN 1293 1499
 FT MOD RES 427 427
 FT METAL 1031 1031
 FT METAL 1035 1035
 FT DOMAIN 467 470
 FT CONFLICT 388 388
 SQ SEQUENCE 1499 AA; 167687 MW; D49964AD635A68D CRC64;
 QY 1 EBYVPMKMSYS 11
 Db 469 EBYVPMKMSYS 479
 Query Match 65.4%; Score 34; DB 1; Length 1499;
 Best Local Similarity 72.7%; Pred. No. 64;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 RESULT 8
 ID RL20 MYCPU STANDARD; PRT; 116 AA.
 AC 0980V0;
 DT 26-FEB-2003 (Rel. 41, Created)
 DT 26-FEB-2003 (Rel. 41, Last sequence update)
 DT 26-FEB-2003 (Rel. 41, Last annotation update)
 DE 50S ribosomal protein L20.
 GN RPL20 OR MYP20 2610.
 OS Mycoplasma pulmonis.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2107;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UAB CTIP.
 RX MEDLINE=21267165; PubMed=11353084;
 RA Chambaud I., Hellig R., Ferris S., Barbe V., Samson D., Galisson F.,
 RA Moszer I., Dybvig K., Wroblewski H., Viart A., Rocha E.P.C.,
 RA Blanchard A.;
 RT "The complete genome sequence of the murine respiratory pathogen
 RT Mycoplasma pulmonis".
 RL Nucleic Acids Res. 29:2145-2153(2001).
 CC -1- FUNCTION: This protein binds directly to 23s ribosomal RNA and is
 CC necessary for the in vitro assembly process of the 50S ribosomal
 CC subunit. It is not involved in the protein synthesizing functions
 CC of that subunit (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE L20P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL: AL445563; CAC1334.1; -.
DR PIR: E90544; B90544.
DR MyPurisc; MYPU_2610; -.
DR HAMAP; MF_00382; -.
DR InterPro; IPR005813; L20.
DR InterPro; IPR005812; L20_bact.org.
DR Pfam; PF00453; RIBOSOMAL_L20_1.
DR PRINTS; PR00062; RIBOSOMAL_L20.
DR ProDom; PD002389; L20; 1.
DR TIGRFAMs; TIGR01032; TP1T_bact; 1.
DR POSITIVE; PS00937; RIBOSOMAL_L20; 1.
KM RIBOSOMAL protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 116 AA; 13565 MW; C59C748901B18F14 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 116;
Best Local Similarity 77.8%; Pred. No. 7.1;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Dy 3 EVVFXGMSYS 11
Db 68 VRPLGMSYS 76

RESULT 9
YJ49_ARCFU STANDARD; PRT; 165 AA.
ID YJ49_ARCFU
AC 028330;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Hypothetical protein AF1949.
DE AF1949.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervazage A.R., Leah N.H., Sutton G.G., Gill S.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Stitt G.G., Gill S.,
RA Kirnhaas B.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB000968; AA899307.1; -.
DR EMBL; D69493; D69493.
DR TIGR; AF1949; -.
KM Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 7 27 POTENTIAL.
SQ SEQUENCE 141 161 POTENTIAL.
SQ 165 AA; 17588 MW; BEC17054810ADB8 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 165;

```

```

Best Local Similarity 60.0%; Pred. No. 10;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Dy 1 EEVFXGMSY 10
Db 60 EESIPDGASY 69

RESULT 10
Y990_CAMJTB STANDARD; PRT; 253 AA.
ID Y990_CAMJTB
AC P45489; Q9PVT0;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein Cj0990C.
GN Cj0990C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Spirochaetobacteriales; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagers K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrett B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
RN [2]
RP SEQUENCE OF 160-253 FROM N.A.
RC STRAIN=ATCC 43431 / TGH 9011;
RX MEDLINE=95247673; PubMed=7730270;
RA Hani B.K., Chan V.L.;
RT "Expression and characterization of Campylobacter jejuni
RT benzoyl-glycine amidohydrolase (Hippuricase) gene in Escherichia
RT coli."
RL J. Bacteriol. 177:2396-2402(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL139076; CAB7246.1; -.
DR EMBL; Z36940; CAA65392.1; -.
DR PIR; C81374; C81374.
DR PIR; I40758; I40758.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 253 AA; 29783 MW; F96D33F73255FA6A CRC64;

Query Match 63.5%; Score 33; DB 1; Length 253;
Best Local Similarity 55.6%; Pred. No. 16;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Dy 2 EVVFXGMSY 10
Db 185 DIPFGMSY 193

RESULT 11
CTX3_MOUSE STANDARD; PRT; 280 AA.
ID CTX3_MOUSE
AC Q9D387; Q9CKQ4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)

```

DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Protein C20orf103 homolog precursor.
 GN C20orf103.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aikawa T., Hara A., Furukoshi Y., Komuro H., Adachi S., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamamoto I.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kawakura T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
 RA Fletschmann W., Gaasterland T., Glasl C., King B., Kochiya H.,
 RA Kuehl P., Lewis S., Matsumoto Y., Nishida I., Pesole G., Quackenbush J.,
 RA Schmitt L.M., Staubli P., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baladrelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bona M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guerninich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Nozaki H., Sato K., Schoenbach C., Seiya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszewski B., Yoshida K., Hasegawa Y., Kawai H., Kontani S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690(2001).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Mammary fibroblast;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Klausner R.D., Collins P.S., Wagner L., Shemen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
 RA Diatchenko L., Marisina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stetler-Stevenson M.G., Seldin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Heaton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitney M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska D.B.,
 RA Scherch A., Schein J.B., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9D387-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9D387-2; Sequence=VSP_003820;
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
 CC in positions 174 and 239.
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL, AK014127; BAB29169.1; ALT_FRAME.
 DR EMBL, AK018222; BAB31124.1; ALT_FRAME.
 DR EMBL, BC004791; AAB04791.1; -.
 DR MGI:1920368; 3110035N03R1K.
 DR MGI:1923411; 6330527006R1K.
 KW Transmembrane; Signal; Alternative splicing.
 FT SIGNAL 1 29
 FT CHAIN 30 280
 FT DOMAIN 30 235
 FT TRANSMEM 236 256
 FT DOMAIN 257 280
 FT CARBOHYD 35 35
 FT CARBOHYD 53 53
 FT CARBOHYD 102 102
 FT CARBOHYD 127 127
 FT VAAPSLIC 1 118
 FT CONFLICT 221 221 E -> V (IN REF. 1; BAB31124).
 FT CONFLICT 230 230 O -> P (IN REF. 1; BAB31124).
 FT CONFLICT 238 238 P -> A (IN REF. 1; BAB31124).
 SQ SEQUENCE 280 AA; 31721 MW; FA1D7BF9PDSCEP CRC64;
 Query Match 63.5%; Score 33; DB 1; Length 280;
 Best Local Similarity 75.0%; Pred. No. 18;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 3 VVPYKMSY 10
 Db 173 VTPAGMSY 180
 RESULT 12
 AROA_VIBCH STANDARD; PRT; 426 AA.
 ID AROA_VIBCH
 AC 09XRB0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
 DB ENOLPYRUVYLKIMATE-3-PHOSPHATE SYNTHASE) (EPSPS).
 GN AROA OR VC1732.
 OS Vibrio cholerae.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 CC Vibrionaceae; Vibrrio.
 CC NCBI_TaxID=666;
 CC [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BI Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Haft D.H., Hickey R.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dracoli I., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 cholerae";
 RL Nature 406:477-483(2000).
 CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
 CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
 CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
 CC sixth step.
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SIMILARITY: Belongs to the EPSP synthase family.
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

CC      or send an email to license@ebi-eb.ch).
CC      -----
DR      EMBL; AE004251; AAP94882.1;  -.
DR      PIR; D82163; D82163.
DR      TIGR; VC1732;  -.
DR      HAMAP; MF_00210; -; 1.
DR      InterPro; IP0006264; ArcoA.
DR      InterPro; IP0001986; BESP_gynthase.
DR      Pfam; PF00275; BESP_gynthase; 1.
DR      ProDom; PD001867; BESP_gynthase; 1.
DR      TIGRFAMs; TIGR01356; arcoA; 1.
DR      PROSITE; PS00104; BESP_SYNTHASE_1; 1.
DR      PROSITE; PS00885; BESP_SYNTHASE_2; 1.
KM      Aromatic amino acid biosynthetasis; Transferrase; Complete proteome.
SO      SEQUENCE 426 AA; 46101 MW; 3865226483BF61C3 CRC64;

Query Match          63.5%; Score 33; DB 1; Length 426;
Best Local Similarity 60.0%; Pred. NO. 28;
Matches      6; Conservative      1; Mismatches      3; Indels      0; Gaps      0;

QY      1  ERYVPXGMSY 10
        |||:|:|
Db       223  EFVIPAGQSY 232

RESULT 13
CC37_SCHPO STANDARD; PRT; 466 AA.
ID   CC37_SCHPO
AC   O94740;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hsp90 co-chaperone Cdc37 (Hsp90 chaperone protein kinase-targeting
DE   subunit) (Cell division control protein 37).
GN   CDC37 OR SPBC9B6.10.
OS   Schizosaccharomyces pombe (fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
ON   [1]
RP   SEQUENCE FROM N.A.
RP   Westwood P.K., Preston N.C., Fantes P.A.;
RT   "Schizosaccharomyces pombe cdc37 gene.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21846401, PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgourou S., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holtrová S., Hornby T., Howarth S., Huckle B.J., Hunt S., Jørgels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., O'Neill S., Mungall K., Murphy L., Niblett D., Odeli C.,
RA   Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
RA   Ruberford K., Rutter S., Saunders D., Seeger S., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Wolkeste G., Aert R., Robben J., Grymourez B.,
RA   Weljens J., Valtersels B., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritz C., Hiltz C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Leclaire V., Mottier S.,
RA   Galibert F., Aves S.-J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerniuk L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

```

RT "The genome sequence of Schistosaccharomyces pombe."
RL Accession: F51871-880(2002).
CC -1- FUNCTION: With Hsp90 it forms a complex that binds to several
CC kinases, resulting in stabilization and promotion of their
CC activity (By similarity).
CC -1- SUBUNIT: Forms a complex with Hsp90. Interacts with a number of
CC kinases (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE CDC37 FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/announce/
CC or_send_an_email_to_license@isb-sib.ch](http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch)).

DR EMBL: AJ132377; CAB38758.1; -
DR EMBL: AJ132376; CAB38757.1; -
DR EMBL: AL049769; CAB42371.2; -
DR PIR: T43653; T43653.
DR GeneDB: SPombe; SPBC9B.10; -
DR InterPro: IPR004918; Cdc37.
DR Pfam: PF03234; Cdc37; 1.
DR Chromosome: Cell division; Cell cycle.
SQ SEQUENCE 466 AA; 52554 MW; 647238B34CABB3C5 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 466;
Best Local Similarity 50.0%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVPEKGMXY 10
DB : |||||
98 DSAIPGGMXY 107

RESULT 14
GSR2 HUMAN STANDARD; PRT; 478 AA.
ID GSR2 HUMAN
AC Q9NZM5; Q9BTC6; Q9HAX6; Q9NPP1; Q9NPR4; Q9UFI2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DT G11oma tumor suppressor candidate region gene 2 protein (p60).
DB G11oma tumor suppressor candidate region gene 2 protein (p60).
GN G11SCR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
OX [1]
RN SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.
RX MEDLINE=20175430; PubMed=10708517;
RA Smith J.S., Tachibana I., Pohl U., Lee H.K., Thamarajasingam U.,
RA Pottier B.P., Ueki K., Billings S., Ramaswamy S., Motreweiser H.W.,
RA Schweitauer B.W., Louis D.N., Jenkins R.B.,
RA "A transcript map of the chromosome 19q-Arm glioma tumor suppressor
RT region."
RL Genomics 64:44-50(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Muscle;
RC MEDLINE=22388257; PubMed=12477932;
RA Klausner R.L., Feingold B.A., Grouse L.H., Derge J.G.,
RA Altschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Ditschenko L., Maruoka K., Farmer A.A., Rubin G.M., Hong L.,
RA Stempleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.B.,
RA Brownlee M.J., Ueda T.B., Toshitaki S., Cavinich P., Prange C.,
RA Rata S.S., Loguailano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McKean P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.U., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton B., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Bikesley R.W., Touchman V.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.B.,
RA Scherch J., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RL [3]
RN SEQUENCE OF 9-478 FROM N.A., AND SUBCELLULAR LOCATION.
RX MEDLINE=99214318; PubMed=10196275;
RA Bruni R., Fineschi B., Ogle W.O., Roizman B.;
RT "A novel cellular protein, p60, interacting with both herpes simplex
RT virus 1 regulatory proteins ICP22 and ICP0 is modified in a
RT cell-type-specific manner and is recruited to the nucleus after
RT infection.";
RL J. Virol. 73:3810-3817 (1999).
RL [4]
RP SEQUENCE OF 12-478 FROM N.A.
RA Andreu N., Baciivil X., Bercarteller M., Sunmy L.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 218-477 FROM N.A.
RC TISSUE-Testis;
RA Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: Interacts with HSV-1 early proteins ICP22 and ICP0.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Expressed at high levels in heart and
CC pancreas, moderate levels in placenta, liver, skeletal muscle, and
CC kidney, and low levels in brain and lung.
CC -1- SIMILARITY: BELONGS TO THE GLTSCR2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF182076; AAF62873.1; -
DR EMBL; BC004229; AAH04229.1; -
DR EMBL; BC006311; AAH06311.1; -
DR EMBL; BC010095; AAH10095.1; -
DR EMBL; AF296124; AAG30413.1; -
DR EMBL; AL359335; CAB94786.1; -
DR EMBL; AL359336; CAB94787.1; -
DR EMBL; AL122063; CAB92424.1; -
DR SWISS-2DPAGE; Q9NZM5; HUMAN.
DR Genew; HGNC:4333; GLTSCR2.
DR MIM; 605691; -
DR GO; GO:0008181; F: tumor suppressor; TAS.
KM Nuclear protein; Polymorphism.
FT VARIANT 389 389
FT R -> Q.
FT /FTID=VAR_011486.
FT GGS -> HEG (IN REF. 2; AAH04229).
FT CONFLICT 4 6
FT CONFLICT 9 9
FT CONFLICT 146 191
FT TVERP -> SGRSSYGRWSPRASPFGAQPSPVADPFCN
FT KQNPAPGHRIDA (IN REF. 3).
FT SDNPLDRPLVQDDEFFLE -> LNNPDKPVVWPGCLFPG
FT (IN REF. 3).
FT A -> S (IN REF. 3).
FT D -> H (IN REF. 2; AAH04229).
FT D -> H (IN REF. 3).
FT PEGNTLRFRFSFGRRNRIIRPERAKFKRKYKLVKVERAR
FT RRIQ -> VLVTSCKGACPVWTBLLFVPRGGRHRC
FT WAGPVPMPRG (IN REF. 5).
FT EGNILDRFRFSFORNMIEPRERAKFKRKYKLVKVERAFR
FT EGIOL -> RGOHSFETGSRAPRGCI (IN REF. 3).
FT SEQUENCE 478 AA; 54417 MW; 7F18923B348CB52B CRC64;

Query Match 63.5%; Score 33; DB 1; Length 478;
Best Local Similarity 60.0%; Pred. No. 31;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Dy 2 EVVEXGMSYS 11
Db 239 EVAPAGASYN 248
RESULT 15
BCNS_CLOPE
ID BCNS_CLOPE STANDARD; PRT; 890 AA.
AC P08696;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DB Bacteriocin BCNS.
GN BCN.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CPN50.
RX MEDLINE=88356297; PubMed=2901768;
RA Garnier T., Cole S.T.;
RT "Complete nucleotide sequence and genetic organization of the
RT bacteriocinogenic plasmid, pIP404, from Clostridium perfringens.";
RL Plasmid 19:134-150 (1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CPN50;
RX MEDLINE=87057020; PubMed=2877971;
RA Garnier T., Cole S.T.;
RT "Characterization of a bacteriocinogenic plasmid from Clostridium
RT perfringens and molecular genetic analysis of the
RT bacteriocin-encoding gene.";
RL J. Bacteriol. 168:1189-1196 (1986).
RN [3]
RP SEQUENCE OF 1-14 FROM N.A.
RC STRAIN=CPN50.
RX MEDLINE=89039249; PubMed=2460717;
RA Garnier T., Cole S.T.;
RT "Studies of UV-inducible promoters from Clostridium perfringens in
RT vivo and in vitro";
RL Mol. Microbiol. 2:607-614 (1988).
CC -1- FUNCTION: MAY FUNCTION AS AN IONOPHORE.
CC -1- INDUCTION: By UV irradiation.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M14481; AAA98248.1; -
DR EMBL; M32882; AAA98249.1; -
DR PIR; A30481; A30481.
DR InterPro; IPR003646; SH3_bac.
DR Pfam; PF00246; Zn_cardoepct. 1.
DR SMART; SM00287; SH3b; 3.
KM Antibiotic; Bacteriocin; Plasmid.
FT DOMAIN 815 869 HYDROPHOBIC.
FT SEQUENCE 890 AA; 96699 MW; F4B58971C31C6C6 CRC64;
SO
Query Match 63.5%; Score 33; DB 1; Length 890;
Best Local Similarity 66.7%; Pred. No. 60;

Matches	6;	Conservative	1;	Mismatches	2;	Indels	0;	Gaps	0;
Qy	2	EVVFXGMSY	10						
		:							
Db	170	EVVFGGFTY	178						

Search completed: December 22, 2003, 17:42:30
Job time : 4.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 17:27:26 ; Search time 25.2 Seconds
(without alignments)
112.642 Million cell updates/sec

Title: US-09-909-164-47
Perfect score: 52
Sequence: 1 EEVVPXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_RVIRUS:*
16: SP_BACTERIAD:*
17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	75.0	1044	16	Q8DIH0
2	38	73.1	387	16	Q98FX1
3	38	73.1	3472	1	Q74056
4	37	71.2	840	3	Q9URY8
5	37	71.2	1123	16	Q8BWD4
6	36	69.2	471	11	Q8RI26
7	36	69.2	484	11	Q8VD18
8	36	69.2	484	11	Q8BTX4
9	36	69.2	484	11	Q8BK35
10	35	67.3	225	10	Q40129
11	35	67.3	425	5	Q9XVK4
12	35	67.3	495	11	Q8C1D7
13	35	67.3	556	4	Q43733
14	35	67.3	583	5	Q9BH83
15	35	67.3	583	5	Q9BH45
16	35	67.3	583	5	Q815S7

17	35	67.3	670	11	Q01487	Q01487 rattus rat
18	35	67.3	747	16	Q8PM16	Q8PM16 xanthomonas
19	35	67.3	747	16	Q8PAT2	Q8PAT2 xanthomonas
20	35	67.3	749	16	Q8PDM6	Q8PDM6 xyloella fas
21	35	67.3	1302	4	Q14122	Q14122 homo sapien
22	34	65.4	156	3	Q12479	Q12479 saccharomyc
23	34	65.4	219	17	Q97182	Q97182 sulfolobus
24	34	65.4	252	17	Q28342	Q28342 archaeoglob
25	34	65.4	290	4	Q96M1	Q96M1 homo sapien
26	34	65.4	387	16	Q92MD6	Q92MD6 rhizobium m
27	34	65.4	489	4	Q81YM3	Q81YM3 homo sapien
28	34	65.4	541	16	Q98BP5	Q98BP5 rhizobium 1
29	34	65.4	544	16	Q9PD2	Q9PD2 ureaplasma
30	34	65.4	842	3	Q9URR4	Q9URR4 penicillium
31	34	65.4	899	16	Q8G4T5	Q8G4T5 bifidobacte
32	34	65.4	1049	16	Q8XT05	Q8XT05 ralbionia s
33	33	63.5	143	17	Q8TX62	Q8TX62 methanopyru
34	33	63.5	166	16	Q8PP5	Q8PP5 xanthomonas
35	33	63.5	193	2	Q8VUA8	Q8VUA8 lactococcus
36	33	63.5	208	2	Q8RTQ4	Q8RTQ4 candidatus
37	33	63.5	209	16	Q8RES6	Q8RES6 fusobacteri
38	33	63.5	284	16	P74187	P74187 synecocyst
39	33	63.5	298	10	Q9M3C0	Q9M3C0 arabidopsis
40	33	63.5	326	12	Q9QCE7	Q9QCE7 soil-borne
41	33	63.5	326	12	Q9Q9Q7	Q9Q9Q7 soil-borne
42	33	63.5	326	12	Q9DJG4	Q9DJG4 soil-borne
43	33	63.5	326	12	Q9OQ99	Q9OQ99 soil-borne
44	33	63.5	326	12	Q91DNI	Q91DNI soil-borne
45	33	63.5	326	12	Q9Q9Q5	Q9Q9Q5 soil-borne

ALIGNMENTS

RESULT 1	Q8DIH0	PRELIMINARY;	PRT;	1044 AA.
ID	Q8DIH0			
AC	Q8DIH0			
DT	01-MAR-2003 (TREMBLrel. 23, Created)			
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DE	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
GN	Multidrug efflux transporter.			
GN	TLA1618.			
OS	Synecococcus elongatus (Thermosynechococcus elongatus).			
OC	Bacteria; Cyanobacteria; Chroococcales; Synecococcus.			
OX	NCBI_TaxID=32046;			
RP	NCBI [1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BP-1;			
RX	MEDLINE=22225144; PubMed=12240834;			
RA	Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Saramoto S.,			
RA	Watanabe A., Iriyuchi M., Kawashima K., Kimura T., Kishida Y.,			
RA	Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,			
RA	Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;			
RT	*Complete genome structure of the thermophilic cyanobacterium			
RT	Thermosynechococcus elongatus BP-1.;			
RL	DNA Res. 9:123-130(2002).			
DR	EMBL; AF005374; BAC09170.1; .			
KW	Complete proteome.			
SC	SEQUENCE 1044 AA; 113205 MW; 00B9C13F0F636D2F CRC64;			
QY	Query Match	75.0%;	Score 39;	DB 16; Length 1044;
QY	Best local Similarity	63.6%;	Pred. No. 24;	
QY	Matches	7;	Conservative	2; Mismatches
QY				2; Indels
QY				0; Gaps
QY				0;
QY	1 EEVVPXGMSYS 11			
QY	: :			
QY	843 EEVLPNGICYS 853			
DB				
RESULT 2	Q98FX1	PRELIMINARY;	PRT;	387 AA.
ID	Q98FX1			

AC Q98FX1;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
 DE Hipurate hydrolase.
 GN MLR3583.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Matnabe A., Ideawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Wochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.";
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AF003002; BAB50445.1; -.
 DR InterPro: IPR002933; Peptidase_M20.
 DR Pfam: PF01546; Peptidase_M20; 1.
 KW Hydrolase; Complete proteome.
 SQ SEQUENCE 387 AA; 41180 MW; 131BPF8E64306829 CRC64;

Query Match 73.1%; Score 38; DB 16; Length 387;
 Best Local Similarity 60.0%; Pred. No. 13;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 10
 DB 367 DEAIPIHGMSY 376

RESULT 3
 ID 074056 PRELIMINARY; PRT; 3472 AA.
 AC 074056;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical 367.1 kDa protein.
 OS Cenarchaeum symbiosum.
 OC Archaea; Crenarchaeota; Thermoprotei; Cenarchaeales; Cenarchaeaceae;
 OC Cenarchaeum.
 OX NCBI_TaxID=46770;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B.
 RX MEDLINE=98422450; PubMed=9748430;
 RA Schleper C., Delong E.F., Preston C.M., Feldman R.A., Wu K.Y.,
 RA Swanson R.V.;
 RT "Genomic analysis reveals chromosomal variation in natural populations
 of the uncultured psychrophilic archaeon Cenarchaeum symbiosum.";
 RL J. Bacteriol. 180:5003-5009(1998).
 DR EMBL: AF083072; AAC62699.1; -.
 DR InterPro: IPR000515; BPD_transp.
 DR Pfam: PF00400; WD40; 4.
 DR SMART: SM00320; WD40; 2.
 DR PROSITE: PS00402; BPD_TRANSP_INN_MEMBER; 1.
 KW Hypothetical protein; Repeat; WD repeat.
 SQ SEQUENCE 3472 AA; 367058 MW; 37F80707030F9355 CRC64;

Query Match 73.1%; Score 38; DB 1; Length 3472;
 Best Local Similarity 54.5%; Pred. No. 1.5e+02;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
 DB 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

DB 2294 EDVIRPGISFS 2304

RESULT 4
 ID Q9URY8 PRELIMINARY; PRT; 840 AA.
 AC Q9URY8;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Probable sulfate permease.
 GN SPAC869.05C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972h-;
 RA Hunt C., Aves S., McDougall R.C., Rajandream M.A., Barrell B.G.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL132779; CAB60015.1; -.
 DR GeneDB: SPombe; SPAC869.05C; -.
 DR InterPro: IPR002645; STAS.
 DR InterPro: IPR001902; Sulph_transpc.
 DR Pfam: PF01740; STAS; 1.
 DR Pfam: PR00916; Sulfate_transp; 1.
 DR TIGRPFAM: TIGR00815; SULP; 1.
 DR PROSITE: PS00801; STAS; 1.
 SQ SEQUENCE 840 AA; 93517 MW; ED4833E162B69077 CRC64;

Query Match 71.2%; Score 37; DB 3; Length 840;
 Best Local Similarity 77.8%; Pred. No. 51;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVVPXGMSYS 11
 DB 135 VVPGMSYA 143

RESULT 5
 ID Q8BMD4 PRELIMINARY; PRT; 1123 AA.
 AC Q8BMD4;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE MYB 2560 paralog; 57%.
 GN MYB2710.
 OS Mycoplasma penetrans.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=28227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HF-2.
 RX MEDLINE=22354719; PubMed=12466555;
 RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
 RA Toshiro C., Horino A., Shiba T., Sasaki T., Hattori M.;
 RT "The complete genomic sequence of Mycoplasma penetrans, an
 RT intracellular bacterial pathogen in humans.";
 RL Nucleic Acids Res. 30:5293-5300(2002).
 DR EMBL: AP004171; BAC44062.1; -.
 KW Complete proteome.
 SQ SEQUENCE 1123 AA; 123636 MW; A4D707330E3DB4AC CRC64;

Query Match 71.2%; Score 37; DB 16; Length 1123;
 Best Local Similarity 70.0%; Pred. No. 70;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVVPXGMSYS 11
 DB 658 EVVPMGLSYS 667

RESULT 6
 Q8R126 PRELIMINARY; PRT; 471 AA.
 AC Q8R126; 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Hypothetical 54.5 kDa protein (fragment).
 GN GITSR2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Strausberg R.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC025810; AAH25810.1; -.
 DR MCD; MGI:2154441; Gtscr2.
 KW Hypothetical protein.
 FT NON TER
 SQ SEQUENCE 471 AA; 54506 MW; E0DA685C374A9760 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 471;
 Best Local Similarity 60.0%; Pred. No. 44;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPGXMSYS 11
 Db 226 EVIPAGASYN 235

RESULT 7
 Q8VD18 PRELIMINARY; PRT; 484 AA.
 AC Q8VD18; 01-MAR-2002 (TREMblrel. 20, Created)
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Similar to glioma tumor suppressor candidate region gene 2.
 GN GITSR2 OR AW536441.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Salivary gland;
 RA Strausberg R.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC017637; AAH17637.1; -.
 DR MGD; MGI:2154441; Gtscr2.
 SQ SEQUENCE 484 AA; 55835 MW; BBB45F3B4BE02A36 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
 Best Local Similarity 60.0%; Pred. No. 45;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPGXMSYS 11
 Db 239 EVIPAGASYN 248

RESULT 8
 Q8BTX4 PRELIMINARY; PRT; 484 AA.
 AC Q8BTX4; 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)

DE Similar to glioma tumor suppressor CANDIDATE region gene 2
 DE protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NOD; TISSUE=Thymus;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 DR EMBL; AK088461; BAC40367.1; -.
 SQ SEQUENCE 484 AA; 55806 MW; B3056425B5ECCAD8 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
 Best Local Similarity 60.0%; Pred. No. 45;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPGXMSYS 11
 Db 239 EVIPAGASYN 248

RESULT 9
 Q8BK35 PRELIMINARY; PRT; 484 AA.
 AC Q8BK35; 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Similar to glioma tumor suppressor CANDIDATE region gene 2
 DE protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Pituitary;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 DR EMBL; AK077341; BAC36760.1; -.
 SQ SEQUENCE 484 AA; 55792 MW; EB67949BCBE92D44 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
 Best Local Similarity 60.0%; Pred. No. 45;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPGXMSYS 11
 Db 239 EVIPAGASYN 248

RESULT 10
 Q40129 PRELIMINARY; PRT; 225 AA.
 AC Q40129; 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Hypothetical 25.2 kDa protein precursor.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamiales; Solanales; Solanaceae; Solanum.

OX NCBI_TaxID=4081;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=VF36; TISSUE=Pistil;
RX MEDLINE=9537523; PubMed=7647301;
RA Milligan S.B., Gasser C.S.;
RT "Nature and regulation of pistil-expressed genes in tomato.";
RL Plant Mol. Biol. 28:691-711(1995).
DR EMBL; U20592; AAA80497.1; -
DR InterPro; IPR002160; Kunitz_legume.
DR Pfam; PF00197; Kunitz_legume; 1.
DR PRINTS; PR00291; KUNITZINBTR.
DR ProDom; PD000891; Kunitz_legume; 1.
DR SMART; SM00452; STR; 1.
DR PROSITE; PS00283; SOYBEAN_KUNITZ; 1.
KW Hypothetical protein; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 225 UNKNOWN.
SQ SEQUENCE 225 AA; 25188 MW; 1074C261D20CFDAD CRC64;

Query Match 67.3%; Score 35; DB 10; Length 225;
Best Local Similarity 54.5%; Pred. No. 32;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
DB 32 DEVVPMGKTYA 42

RESULT 11

OXVKA
ID 09XVKA PRELIMINARY; PRT; 425 AA.
AC 09XVKA;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE R10012.10 protein.
GN R10012.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN
RP SEQUENCE FROM N.A.
RA Percy C.M.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81109; CAB03241.1; -
DR WormBep; R10012.10; CE12690.
DR InterPro; IPR000719; Prot_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 425 AA; 49410 MW; 5D96E29B08CE9D6 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 425;
Best Local Similarity 50.0%; Pred. No. 64;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 10
DB 335 EQIVPGGLQY 344

RESULT 12

Q8C1D7

ID 08C1D7 PRELIMINARY; PRT; 495 AA.
AC 08C1D7;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE CDNA FLJ31891 FIS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK028258; BAC25846.1; -
SQ SEQUENCE 495 AA; 55358 MW; B6A22A093C114752 CRC64;

Query Match 67.3%; Score 35; DB 11; Length 495;
Best Local Similarity 63.6%; Pred. No. 75;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
DB 260 EEIVPMGISPS 270

RESULT 13

ID 043733 PRELIMINARY; PRT; 556 AA.
AC 043733;
DT 01-JUN-1998 (T-EMBLrel. 06, Created)
DT 01-JUN-1998 (T-EMBLrel. 06, Last sequence update)
DT 01-JUN-1998 (T-EMBLrel. 06, Last annotation update)
DE DNA binding protein (Fragment).
GN DJ451B15.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN
RP SEQUENCE FROM N.A.
RA Tubby B.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z98050; CAB10847.1; -
FT NON_TER 1 1
SQ SEQUENCE 556 AA; 59059 MW; ECB00E403FB2528 CRC64;

Query Match 67.3%; Score 35; DB 4; Length 556;
Best Local Similarity 66.7%; Pred. No. 86;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
DB 244 VVPAGLTVS 252

RESULT 14

Q9BH83 PRELIMINARY; PRT; 583 AA.
AC Q9BH83;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Choline transporter.
GN SCT1.
OC Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Ben Mamoun C., Gluzman I.Y., Goldberg D.E.;
 RT "Plasmodium falciparum choline transporter (PfSCT1) gene."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY007374; AAK14818.1; -.
 DR EMBL; AY007373; AAK14817.1; -.
 DR InterPro; IPR002123; Acyltransferase.
 DR Pfam; PF01553; Acyltransferase; 1.
 DR SMART; SM00563; Pfam; 1.
 SQ SEQUENCE 583 AA; 66917 MW; 4DF83D7530527474 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 583;
 Best Local Similarity 55.6%; Pred. No. 90;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
 :|||:
 Db 227 IIPVGLSYS 235

RESULT 15

Q9BHA5 PRELIMINARY; PRT; 583 AA.
 AC Q9BHA5;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Choline transporter.
 GN SCT1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ben Mamoun C., Gluzman I.Y., Goldberg D.E.;
 RT "Plasmodium falciparum choline transporter (PfSCT1) gene."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY007372; AAK14816.1; -.
 DR EMBL; AY007375; AAK17947.1; -.
 DR InterPro; IPR002123; Acyltransferase.
 DR Pfam; PF01553; Acyltransferase; 1.
 DR SMART; SM00563; Pfam; 1.
 SQ SEQUENCE 583 AA; 66917 MW; 2B2BFAE3E395E049 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 583;
 Best Local Similarity 55.6%; Pred. No. 90;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
 :|||:
 Db 227 IIPVGLSYS 235

Search completed: December 22, 2003, 17:51:34
 Job time : 26.2667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:41:00 ; Search time 32.4667 Seconds
(without alignments)
53.778 Million cell updates/sec

Title: US-09-909-164-5
Perfect score: 52
Sequence: 1 EEVVPXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	96.2	11	23	ABBB0521 Hepatitis C virus
2	50	96.2	11	23	ABBB0522 Hepatitis C virus
3	50	96.2	11	23	ABBB0525 Hepatitis C virus
4	50	96.2	11	23	ABBB0526 Hepatitis C virus
5	50	96.2	11	23	ABBB0559 Hepatitis C virus
6	50	96.2	11	23	ABBB0563 Hepatitis C virus
7	50	96.2	11	23	ABBB0564 Hepatitis C virus
8	50	96.2	11	23	ABBB0565 Hepatitis C virus
9	50	96.2	11	23	ABBB0566 Hepatitis C virus

10	50	96.2	11	23	ABBB0567 Hepatitis C virus
11	50	96.2	11	23	ABBB0568 Hepatitis C virus
12	46	88.5	11	23	ABBB0524 Hepatitis C virus
13	46	88.5	11	23	ABBB0528 Hepatitis C virus
14	46	88.5	11	23	ABBB0529 Hepatitis C virus
15	46	88.5	11	23	ABBB0561 Hepatitis C virus
16	46	88.5	11	23	ABBB0562 Hepatitis C virus
17	45	86.5	11	23	ABBB0523 Hepatitis C virus
18	45	86.5	11	23	ABBB0527 Hepatitis C virus
19	45	86.5	11	23	ABBB0535 Hepatitis C virus
20	45	86.5	11	23	ABBB0536 Hepatitis C virus
21	45	86.5	11	23	ABBB0539 Hepatitis C virus
22	45	86.5	11	23	ABBB0540 Hepatitis C virus
23	45	86.5	11	23	ABBB0558 Hepatitis C virus
24	45	86.5	11	23	ABBB0560 Hepatitis C virus
25	44	84.6	11	23	ABBB0544 Hepatitis C virus
26	44	84.6	11	23	ABBB0545 Hepatitis C virus
27	44	84.6	11	23	ABBB0549 Hepatitis C virus
28	44	84.6	11	23	ABBB0552 Hepatitis C virus
29	44	84.6	11	23	ABBB0553 Hepatitis C virus
30	42	80.8	11	23	ABBB0530 Hepatitis C virus
31	41	78.8	11	23	ABBB0538 Hepatitis C virus
32	41	78.8	11	23	ABBB0542 Hepatitis C virus
33	41	78.8	11	23	ABBB0543 Hepatitis C virus
34	40	76.9	11	23	ABBB0537 Hepatitis C virus
35	40	76.9	11	23	ABBB0541 Hepatitis C virus
36	40	76.9	11	23	ABBB0547 Hepatitis C virus
37	40	76.9	11	23	ABBB0548 Hepatitis C virus
38	40	76.9	11	23	ABBB0551 Hepatitis C virus
39	40	76.9	11	23	ABBB0556 Hepatitis C virus
40	40	76.9	11	23	ABBB0557 Hepatitis C virus
41	40	76.9	20	20	AAU76810 Hepatitis C virus
42	40	76.9	1022	22	ABG03621 Novel human diagno
43	40	76.9	1022	22	ABG03826 Novel human diagno
44	40	76.9	1022	22	ABG08173 Novel human diagno
45	39	75.0	11	23	ABBB0546 Hepatitis C virus

ALIGNMENTS

RESULT 1					
ABBB0521					
ID	ABBB0521	standard; peptide; 11 AA.			
XX					
AC	ABBB0521;				
XX					
DT	08-OCT-2002	(first entry)			
XX					
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.				
XX					
KW	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.				
KM					
XX					
OS	Synthetic.				
XX					
FH					
FT	Key	Location/Qualifiers			
FT	Modified-site	1 /note= "N-terminal acetyl"			
FT	Modified-site	6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"			
FT					
FT	Modified-site	11 /note= "C-terminal amide"			
FT					
PN	MO200208251-A2.				
XX					
PD	31-JAN-2002.				
XX					
PF	19-JUL-2001; 2001WO-US23169.				
XX					
PR	21-JUL-2000; 2000US-220101P.				
XX					

PA (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
XX
XX virus protease
XX
PS Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
DB 1 EEVVPXGMSYS 11

RESULT 2
ABB80522
ID ABB80522 standard; peptide; 11 AA.
AC ABB80522;
XX
XX 08-OCT-2002 (first entry)
DT
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.
DE
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM
XX
XX virucide.
OS
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C

PT virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
PS
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
DB 1 EEVVPXGMSYS 11

RESULT 3
ABB80525
ID ABB80525 standard; peptide; 11 AA.
AC ABB80525;
XX
XX 08-OCT-2002 (first entry)
DT
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #5.
DE
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM
XX
XX virucide.
OS
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
XX
XX virus protease
XX
PS Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of

CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

CC Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
|||
Db 1 EEVVPXGMSYS 11

RESULT 4

ABB80526
ID ABB80526 standard; peptide; 11 AA.

AC ABB80526;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #6.

KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM virucide.

OS Synthetic.

FT Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "N-terminyl carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

PN WO200208251-A2.

PD 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

PA (CORV-) CORVAS INT INC.

PI Lim-wilby M, Levy OE, Brunck TK;

PS WPI; 2002-361643/39.

PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
PS Claim 17; Page 64; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

CC Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
|||
Db 1 EEVVPXGMSYS 11

RESULT 5

ABB80559
ID ABB80559 standard; peptide; 11 AA.

AC ABB80559;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #39.

KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM virucide.

OS Synthetic.

FT Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "N-terminyl carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Modified-site 8 /note= "Oxymethionine"

FT Modified-site 11 /note= "C-terminal amide"

PN WO200208251-A2.

PD 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

PA (CORV-) CORVAS INT INC.

PI Lim-wilby M, Levy OE, Brunck TK;

PS WPI; 2002-361643/39.

PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
PS Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

CC Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
|||

Db	1	EEVVPXGMSYS	11	
RESULT 6				
ABB80563				
ID	ABB80563	standard; peptide; 11 AA.		
XX				
AC	ABB80563;			
XX				
DT	08-OCT-2002	(first entry)		
XX				
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #43.			
XX				
KM	Hepatitis C virus; HCV, serine protease; inhibitor; alpha-ketoamide; virucide.			
XX				
OS	Synthetic.			
XX				
FH	Key	Location/Qualifiers		
FT	Modified-site	1		
FT		/note= "N-terminal acetyl"		
FT	Modified-site	6		
FT		/note= "Valyl carboxyl forming keto-amide linkage with residue 7"		
FT	Modified-site	11		
FT		/note= "C-terminal amide"		
XX				
PN	WO200208251-A2.			
XX				
PD	31-JAN-2002.			
XX				
PF	19-JUL-2001; 2001WO-US23169.			
XX				
PR	21-JUL-2000; 2000US-220101P.			
XX				
PA	(CORV-) CORVAS INT INC.			
XX				
P1	Lim-wilby M, Levy OE, Brunck TK;			
XX				
DR	WPI; 2002-361643/39.			
XX				
PT	Novel, peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C			
PT	virus protease -			
XX				
PS	Claim 17; Page 65; 69pp; English.			
XX				
CC	The sequence represents a peptide compound of the invention having			
CC	hepatitis C virus (HCV) protease inhibitory activity. The peptides of the			
CC	invention are alpha-ketoamide peptide analogues. The peptides have			
CC	virucide activity, and are useful for treating and in the manufacture of			
CC	a pharmaceutical to treat disorders associated with HCV protease. A			
CC	pharmaceutical composition comprising the peptide as an active ingredient			
CC	is useful for treating disorders associated with hepatitis C virus.			
XX				
SQ	Sequence	11 AA;		
Query Match		96.2%; Score 50; DB 23; Length 11;		
Best Local Similarity		100.0%; Pred. No. 0.0013;		
Matches	11; Conservative	0; Mismatches	0; Indels	0; Gaps
OY	1	EEVVPXGMSYS	11	
DB	1	EEVVPXGMSYS	11	
RESULT 7				
ABB80564				
ID	ABB80564	standard; peptide; 11 AA.		
XX				
AC	ABB80564;			
XX				
DT	08-OCT-2002	(first entry)		

XX	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #44.
DE	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX	vitrucide.
KW	Synthetic.
XX	
OS	
XX	
FH	Key
FT	Modified-site Location/Qualifiers 1 /note= "N-terminal acetyl"
FT	Modified-site 6 /note= "Leucyl carbonyl forming keto-amide linkage with residue 7"
FT	Modified-site 11 /note= "C-terminal amide"
PN	WO200208251-A2.
PD	31-JAN-2002.
XX	
PF	19-JUL-2001; 2001WO-US23169.
XX	
PR	21-JUL-2000; 2000US-220101P.
XX	(CORV-) CORVAS INT INC.
PA	
PI	Lim-wilby M, Levy OE, Brunck TK;
XX	WPI; 2002-361643/39.
DR	
XX	
PT	Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease -
PS	Claim 17; Page 65; 6pp; English.
XX	
CC	The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have vitrucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus.
CC	
CX	
SC	Sequence 11 AA:
Query Match	Score 2%; Score 50; DB 23; Length 11; Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative	0; Mismatches 0; Indels 0; Gaps 0.
OY	1 EEVVPXGMSYS 11 Db 1 EEVVPXGMSYS 11
RESULT 8	
ID	ABB80565 standard; peptide; 11 AA.
AC	ABB80565;
DT	08-OCT-2002 (first entry)
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #45.
XX	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM	vitrucide.
XX	
OS	Synthetic.
XX	
FH	Key
FT	Modified-site Location/Qualifiers 1

FT	/note= "N-terminal acetyl"
PT	Modified-site
FT	6 /note= "Norleucyl carbonyl forming keto-amide linkage with residue 7"
FT	Modified-site
FT	11 /note= "C-terminal amide"
XX	WO200208251-A2.
PX	
PD	31-JAN-2002.
XX	
XX	19-JUL-2001; 2001WO-US23169.
PF	
PR	21-JUL-2000; 2000US-220101P.
XX	
PA	(CORV-) CORVAS INT INC.
XX	
PI	Lim-wilby M, Levy OE, Brunck TK;
DR	WPI; 2002-361643/39.
XX	
PT	Novel peptide compound having hepatitis C virus protease inhibitory
PT	activity useful for treating disorders associated with hepatitis C
PT	virus protease -
PS	Claim 17; Page 65; 69pp; English.
XX	
CC	The sequence represents a peptide compound of the invention having
CC	hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC	invention are alpha-ketamide peptide analogues. The peptides have
CC	vinucleic activity, and are useful for treating and in the manufacture of
CC	a medicament to treat disorders associated with HCV protease. A
CC	pharmaceutical composition comprising the peptide as an active ingredient
CC	is useful for treating disorders associated with hepatitis C virus.
XX	
SQ	Sequence 11 AA;
OY	
DB	Query Match 96.2%; Score 50; DB 23; Length 11; Best Local Similarity 100.0%; Pred. No. 0.0013; Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 EEVVPXGMSYS 11 1 EEWVPXGMSYS 11	
RESULT 9	
ABB80566	
ID	ABB80566 standard; peptide; 11 AA.
XX	
AC	ABB80566;
XX	
DT	08-OCT-2002 (first entry)
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #46.
XX	
KW	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketamide;
XX	vinucleic.
OS	Synthetic.
XX	
FH	Key Location/Qualifiers
FT	1 Modified-site
FT	/note= "N-terminal acetyl"
FT	6 Modified-site
FT	/note= "2-aminoisobutyryl carbonyl residue forming a keto-amide linkage with residue 7"
FT	11 Modified-site
FT	/note= "C-terminal amide"
XX	
PN	WO200208251-A2.
XX	
DD	31-JAN-2002.

P	PF	19-JUL-2001; 2001WO-US23169.
X	XX	
P	PR	21-JUL-2000; 2000US-220101P.
X	XX	
P	PA	(CORV-) CORVAS INT INC.
X	XX	
P	PI	Llm-wilby M, Levy OE, Brunck TK;
X	XX	
D	DR	WPI; 2002-361643/39.
X	XX	
P	PT	Novel peptide compound having hepatitis C virus protease inhibitory
X	XX	activity useful for treating disorders associated with hepatitis C
P	PT	virus protease -
X	XX	
P	PS	Claim 17; Page 65; 69pp; English.
X	XX	
C	CC	The sequence represents a peptide compound of the invention having
C	CC	hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
C	CC	invention are alpha-ketamide peptide analogues. The peptides have
C	CC	virtic acid activity, and are useful for treating and in the manufacture of
C	CC	a medicament to treat disorders associated with HCV protease. A
C	CC	pharmaceutical composition comprising the peptide as an active ingredient
C	CC	is useful for treating disorders associated with hepatitis C virus.
X	XX	
SQ	Sequence	11 AA:
Query Match		96.2%; Score 50; DB 23; Length 11;
Best Local Similarity		100.0%; Pred.No. 0.0013;
Matches	11; Conservative	0; Mismatches 0; Indels 0; Gaps 0
OY	1 EEVVPXGMSYS 11 	
Db	1 EEVVPXGMSYS 11	
RESULT 10		
ABB80567		
ID	ABB80567 standard; peptide; 11 AA.	
AC	ABB80567;	
DT	08-OCT-2002 (first entry)	
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #47.	
KM	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.	
OS	Synthetic.	
PH	Key	Location/Qualifiers
FT	Modified-site	1 /note= "N-terminal acetyl"
FT	Modified-site	6 /note= "(S,S) allochreonyl carbonyl residue forming a keto-amide linkage with residue 7"
FT	Modified-site	11 /note= "C-terminal amide"
MO200208251-A2.		
31-JAN-2002.		
19-JUL-2001; 2001WO-US23169.		
21-JUL-2000; 2000US-220101P.		
(CORV-) CORVAS INT INC.		
Llm-wilby M, Levy OE, Brunck TK;		
WPI; 2002-361643/39.		

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX
PS Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013; Mismatches 0; Indels 0; Gaps 0;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
|||
DB 1 EEVVPXGMSYS 11

RESULT 11
ABB80568
ID ABB80568 standard; peptide; 11 AA.
XX
XX ABB80568;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #48.
DE
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM virucide.
XX
XX Synthetic.
OS
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Alpha-propionyl-glycyl-carbonyl residue forming
FT a keto-amide linkage with residue 7"
FT Modified-site 11 /note= "C-terminal amide"
FT
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013; Mismatches 0; Indels 0; Gaps 0;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
|||
DB 1 EEVVPXGMSYS 11

RESULT 12
ABB80524
ID ABB80524 standard; peptide; 11 AA.
XX
XX ABB80524;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.
DE
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM virucide.
XX
XX Synthetic.
OS
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "N-terminal carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;

Query Match 88.5%; Score 46; DB 23; Length 11;

CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
XX
SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013; Mismatches 0; Indels 0; Gaps 0;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
|||
DB 1 EEVVPXGMSYS 11

RESULT 12
ABB80524
ID ABB80524 standard; peptide; 11 AA.
XX
XX ABB80524;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.
DE
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM virucide.
XX
XX Synthetic.
OS
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "N-terminal carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;

Query Match 88.5%; Score 46; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.0088;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
DB 1 EEVVPXGMDYS 11

RESULT 13

ABB80528
ID ABB80528 standard; peptide; 11 AA.

AC ABB80528;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #8.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketamide;
virocide.

XX Synthetic.

FT Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

FT WO200208251-A2.

PD 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

PA (CORV-) CORVAS INT INC.

PI Lim-wilby M, Levy OE, Brunck TK;

PS WPI; 2002-361643/39.

DR Novel peptide compound having hepatitis C virus protease inhibitory

PT activity useful for treating disorders associated with hepatitis C

PS virus protease -

XX Claim 17; Page 64; 69pp; English.

CC The sequence represents a peptide compound of the invention having

CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

CC invention are alpha-ketamide peptide analogues. The peptides have

CC virucide activity, and are useful for treating and in the manufacture of

CC a medicament to treat disorders associated with HCV protease. A

CC pharmaceutical composition comprising the peptide as an active ingredient

CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 88.5%; Score 46; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.0088; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
DB 1 EEVVPXGMDYS 11

RESULT 14
ABB80529
ID ABB80529 standard; peptide; 11 AA.

AC ABB80529;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketamide;
virocide.

XX Synthetic.

FT Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

FT WO200208251-A2.

PD 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

PA (CORV-) CORVAS INT INC.

PI Lim-wilby M, Levy OE, Brunck TK;

PS WPI; 2002-361643/39.

DR Novel peptide compound having hepatitis C virus protease inhibitory

PT activity useful for treating disorders associated with hepatitis C

PS virus protease -

XX Claim 17; Page 64; 69pp; English.

CC The sequence represents a peptide compound of the invention having

CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

CC invention are alpha-ketamide peptide analogues. The peptides have

CC virucide activity, and are useful for treating and in the manufacture of

CC a medicament to treat disorders associated with HCV protease. A

CC pharmaceutical composition comprising the peptide as an active ingredient

CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 88.5%; Score 46; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.0088; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
DB 1 EEVVPXGMDYS 11

RESULT 15

ABB80561
ID ABB80561 standard; peptide; 11 AA.

AC ABB80561;

XX

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 16:37:03 ; Search time 10.1333 Seconds
(without alignments)
45.930 Million cell updates/sec

Title: US-09-909-164-5
Perfect score: 52
Sequence: 1 BEVVPXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued_Patents_AA:*
2: /cgn2_6/pdata/1/1aa/5A_COMB.pep:*
3: /cgn2_6/pdata/1/1aa/5B_COMB.pep:*
4: /cgn2_6/pdata/1/1aa/6A_COMB.pep:*
5: /cgn2_6/pdata/1/1aa/6B_COMB.pep:*
6: /cgn2_6/pdata/1/1aa/6CTUS_COMB.pep:*
7: /cgn2_6/pdata/1/1aa/6CTUS_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	65.4	947	4	US-09-228-986-73 Sequence 73, Appl
2	34	65.4	1407	4	US-09-328-352-7885 Sequence 7885, Ap
3	33	63.5	45	2	US-08-637-759B-236 Sequence 236, App
4	33	63.5	45	3	US-08-871-355A-236 Sequence 236, App
5	33	63.5	45	4	US-09-201-945-236 Sequence 236, App
6	33	63.5	45	6	517197-51 Patent No. 517197
7	33	63.5	410	6	517197-1 Patent No. 517197
8	33	63.5	1394	6	517197-30 Patent No. 517197
9	32	61.5	10	4	US-09-357-952-66 Sequence 66, Appl
10	32	61.5	10	4	US-09-521-650-66 Sequence 66, Appl
11	32	61.5	10	4	US-09-168-888-66 Sequence 66, Appl
12	32	61.5	102	2	US-08-580-988A-23 Sequence 23, Appl
13	32	61.5	152	2	US-08-460-694-4 Sequence 4, Appl
14	32	61.5	152	3	US-08-460-744-4 Sequence 4, Appl
15	32	61.5	152	3	US-07-667-711B-4 Sequence 4, Appl
16	32	61.5	173	1	US-08-193-977-7 Sequence 7, Appl
17	32	61.5	189	2	US-08-464-517-21 Sequence 21, Appl
18	32	61.5	189	2	US-08-246-361A-21 Sequence 21, Appl
19	32	61.5	189	2	US-08-463-772-21 Sequence 21, Appl
20	32	61.5	189	5	PCR-US93-05000-21 Sequence 21, Appl
21	32	61.5	236	2	US-08-464-517-22 Sequence 22, Appl
22	32	61.5	236	2	US-08-246-361A-22 Sequence 22, Appl
23	32	61.5	236	3	US-08-463-772-22 Sequence 22, Appl
24	32	61.5	236	5	PCR-US93-05000-22 Sequence 22, Appl
25	32	61.5	280	3	US-08-464-517-6 Sequence 6, Appl
26	32	61.5	280	3	US-08-463-772-6 Sequence 6, Appl
27	32	61.5	289	2	US-08-246-361A-4 Sequence 4, Appl

28	32	61.5	289	5	PCR-US93-05000-4 Sequence 4, Appl
29	32	61.5	291	5	PCR-US93-05000-6 Sequence 6, Appl
30	32	61.5	292	2	US-08-464-517-23 Sequence 23, Appl
31	32	61.5	292	2	US-08-246-361A-5 Sequence 6, Appl
32	32	61.5	292	2	US-08-246-361A-23 Sequence 23, Appl
33	32	61.5	292	3	US-08-463-772-23 Sequence 23, Appl
34	32	61.5	292	5	PCR-US93-05000-23 Sequence 23, Appl
35	32	61.5	295	1	US-07-947-120-8 Sequence 8, Appl
36	32	61.5	295	1	US-08-472-893A-8 Sequence 8, Appl
37	32	61.5	295	2	US-08-460-694-2 Sequence 2, Appl
38	32	61.5	295	2	US-08-464-517-19 Sequence 19, Appl
39	32	61.5	295	2	US-08-464-517-20 Sequence 20, Appl
40	32	61.5	295	2	US-08-246-361A-19 Sequence 19, Appl
41	32	61.5	295	2	US-08-246-361A-20 Sequence 20, Appl
42	32	61.5	295	3	US-08-463-772-19 Sequence 19, Appl
43	32	61.5	295	3	US-08-463-772-20 Sequence 20, Appl
44	32	61.5	295	3	US-08-460-744-2 Sequence 2, Appl
45	32	61.5	295	3	US-07-667-711B-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-228-986-73
; Sequence 73, Application US/09228986
; Patent No. 6359198
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Neuenhulzen, Niels
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; TITLE OF INVENTION: And Their Use in the Modification of Plant Cell Signalling
; FILE REFERENCE: 11000/1020
; CURRENT APPLICATION NUMBER: US/09/228,986
; CURRENT FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 73
; LENGTH: 947
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-228-986-73

Query Match
Best Local Similarity 65.4%; Score 34; DB 4; Length 947;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 3 VVPXGMSYS 11
Db 686 VMPXGMSYS 694

RESULT 2
US-09-328-352-7885
; Sequence 7885, Application US/09328352
; Patent No. 662958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-039A
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7885
; LENGTH: 1407
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7885

Query Match
Best Local Similarity 65.4%; Score 34; DB 4; Length 1407;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
Db 596 EVPEGLSF 604

RESULT 3

US-08-637-759B-236
; Sequence 236, Application US/08637759B
; Patent No. 5876931
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,759B
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-637-759B-236

Query Match 63.5%; Score 33; DB 2; Length 45;
Best Local Similarity 60.0%; Pred. No. 6.6;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVPXGMSY 10
Db 1 ERIISPLGMSY 10

RESULT 4

US-08-871-355A-236
; Sequence 236, Application US/08871355A
; Patent No. 6015669
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871,355A
; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-871-355A-236

CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/871,355A

FILING DATE: 09-JUN-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB95/02875

FILING DATE: 11-DEC-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Pabst, Patrea L.

REGISTRATION NUMBER: 31,284

REFERENCE/DOCKET NUMBER: RPMS 101 CON

TELECOMMUNICATION INFORMATION:

TELEPHONE: (404) 873-8794

TELEFAX: (404) 873-8795

INFORMATION FOR SEQ ID NO: 236:

SEQUENCE CHARACTERISTICS:

LENGTH: 45 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

US-08-871-355A-236

Query Match 63.5%; Score 33; DB 3; Length 45;
Best Local Similarity 60.0%; Pred. No. 6.6;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVPXGMSY 10
Db 1 ERIISPLGMSY 10

RESULT 5

US-09-201-945-236
; Sequence 236, Application US/09201945
; Patent No. 6342215
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/201,945
; FILING DATE: 08/03/2001
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/637,759
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:

NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 236:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: protein
HYPOTHETICAL: NO
US-09-201-945-236

Query Match 63.5%; Score 33; DB 4; Length 45;
Best Local Similarity 60.0%; Pred. No. 6.6;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
DB 1 EEISPLGWSY 10

RESULT 6
5177197-51
PATENT NO. 5177197
APPLICANT: KANZAKI, TETSUTO;OLOFSSON, ANDERS;MOREN, ANITA;
WERNSTEDT, CHRISTER;HELLMAN, ULF;MIYAZONO, KOHEI;CLAESSON-WELSH,
LENA;HELDIN, CARL-HENRIK
TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
HUMAN TRANSFORMING GROWTH FACTOR-BETA1-BINDING PROTEIN
NUMBER OF SEQUENCES: 53
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/487,343
FILING DATE: 27-FEB-1990
SEQ ID NO: 51
LENGTH: 65
5177197-51

Query Match 63.5%; Score 33; DB 6; Length 65;
Best Local Similarity 45.5%; Pred. No. 10;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 11
DB 52 KEICPGMGYT 62

RESULT 7
5177197-1
PATENT NO. 5177197
APPLICANT: KANZAKI, TETSUTO;OLOFSSON, ANDERS;MOREN, ANITA;
WERNSTEDT, CHRISTER;HELLMAN, ULF;MIYAZONO, KOHEI;CLAESSON-WELSH,
LENA;HELDIN, CARL-HENRIK
TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
HUMAN TRANSFORMING GROWTH FACTOR-BETA1-BINDING PROTEIN
NUMBER OF SEQUENCES: 53
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/487,343
FILING DATE: 27-FEB-1990
SEQ ID NO: 1
LENGTH: 410
5177197-1

Query Match 63.5%; Score 33; DB 6; Length 410;
Best Local Similarity 45.5%; Pred. No. 81;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 11
DB 399 KEICPGMGYT 409

RESULT 8
5177197-30
PATENT NO. 5177197
APPLICANT: KANZAKI, TETSUTO;OLOFSSON, ANDERS;MOREN, ANITA;
WERNSTEDT, CHRISTER;HELLMAN, ULF;MIYAZONO, KOHEI;CLAESSON-WELSH,
LENA;HELDIN, CARL-HENRIK
TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
HUMAN TRANSFORMING GROWTH FACTOR-BETA1-BINDING PROTEIN
NUMBER OF SEQUENCES: 53
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/487,343
FILING DATE: 27-FEB-1990
SEQ ID NO: 30
LENGTH: 1394
5177197-30

Query Match 63.5%; Score 33; DB 6; Length 1394;
Best Local Similarity 45.5%; Pred. No. 3.3e+02;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 11
DB 399 KEICPGMGYT 409

RESULT 9
US-09-357-952-66
SEQUENCE 66, Application US/09357952
PATENT NO. 6248904
GENERAL INFORMATION:
APPLICANT: Zhang, Han-Zhong
APPLICANT: Cai, Sui Xiong
APPLICANT: Drewe, John A.
APPLICANT: Yang, Wu
TITLE OF INVENTION: Fluorescence Screening Assays for Caspases, Peptidases, Protease
TITLE OF INVENTION: Other Enzymes and the Use Thereof
FILE REFERENCE: 1735.0030001
CURRENT APPLICATION NUMBER: US/09/357,952
CURRENT FILING DATE: 1999-07-21
EARLIER APPLICATION NUMBER: US 60/093,642
EARLIER FILING DATE: 21-JUL-1998
NUMBER OF SEQ ID NOS: 139
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 66
LENGTH: 10
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-357-952-66

Query Match 61.5%; Score 32; DB 3; Length 10;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
DB 1 DDIVPCMSY 10

RESULT 10
US-09-521-650-66
SEQUENCE 66, Application US/09521650
PATENT NO. 6335429
GENERAL INFORMATION:
APPLICANT: Weber, Eckard
APPLICANT: Cai, Sui Xiong
APPLICANT: Keana, John F.W.
APPLICANT: Drewe, John A.

APPLICANT: Zhang, Han-Zhong
TITLE OF INVENTION: No. 635429e1 Fluorogenic or Fluorescent Reporter Molecules and
TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence
TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the
FILE REFERENCE: 1735.0290002
CURRENT APPLICATION NUMBER: US/09/521,650
EARLIER FILING DATE: 2000-03-08
EARLIER APPLICATION NUMBER: 09/168,888
EARLIER FILING DATE: 1998-10-09
EARLIER APPLICATION NUMBER: US 60/061,582
EARLIER FILING DATE: 1997-10-10
EARLIER APPLICATION NUMBER: US 09/033,661
NUMBER OF SEQ ID NOS: 142
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 66
LENGTH: 10
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:Synthetic
US-09-521-650-66

Query Match 61.5%; Score 32; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
DB 1 DDIVPCSMY 10

RESULT 11
US-09-168-888-66
Sequence 66, Application US/09168888
Patent No. 6342611
GENERAL INFORMATION:
APPLICANT: Weber, Eckard
APPLICANT: Cai, Sui Xiong
APPLICANT: Keana, John F.W.
APPLICANT: Drewe, John A.
APPLICANT: Zhang, Han-Zhong
TITLE OF INVENTION: No. 634261e1 Fluorogenic or Fluorescent Reporter Molecules and
TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence
TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the
FILE REFERENCE: 1735.0290002
CURRENT APPLICATION NUMBER: US/09/168,888
CURRENT FILING DATE: 1998-10-09
EARLIER APPLICATION NUMBER: US 60/061,582
EARLIER FILING DATE: 1997-10-10
EARLIER APPLICATION NUMBER: US 09/033,661
EARLIER FILING DATE: 1998-03-03
NUMBER OF SEQ ID NOS: 142
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 66
LENGTH: 10
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:Synthetic
US-09-168-888-66

Query Match 61.5%; Score 32; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 1 EEVVPXGMSY 10
DB 1 DDIVPCSMY 10

RESULT 12
US-08-580-988A-23
Sequence 23, Application US/08580988A
Patent No. 5856161
GENERAL INFORMATION:
APPLICANT: Agarwal et al.
TITLE OF INVENTION: Tumor Necrosis Factor
TITLE OF INVENTION: Receptor-1-Associated Protein Kinase And Methods
TITLE OF INVENTION: For Its Use
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dr. Benjamin A. Adler
STREET: 8011 Candle Lane
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77071

COMPUTER READABLE FORM:
MEDIUM TYPE: 1.44 Mb floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word for Macintosh
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/580,988A
FILING DATE: January 3, 1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Benjamin Aaron Adler, Ph.D., J.D.
REGISTRATION NUMBER: 35,423
REFERENCE/DOCKET NUMBER: D5721CIP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-777-2321
TELEFAX: 713-777-6908

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:
LENGTH: 102 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: Protein
HYPOTHETICAL: no
ANTI-SENSE: no
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
US-08-580-988A-23

Query Match 61.5%; Score 32; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
DB 24 EEVPLAMNY 33

RESULT 13
US-08-460-694-4
Sequence 4, Application US/08460694
Patent No. 5858655
GENERAL INFORMATION:
APPLICANT: Arnold, Andrew
TITLE OF INVENTION: PRAD1 Cyclin and its cDNA
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington

Query Match 61.5%; Score 32; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 1 EEVVPXGMSY 10
DB 1 DDIVPCSMY 10

```

; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,694
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McConathy, Evelyn H.
; REGISTRATION NUMBER: 35,279
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 152 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-460-694-4

Query Match      61.5%; Score 32; DB 2; Length 152;
Best Local Similarity 60.0%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      1 EEVVPXGMSY 10
        ||| | | |
        20 EEVFPPLMNY 29

RESULT 14
US-08-460-744-4
; Sequence 4, Application US/08460744
; Patent No. 6107541
; GENERAL INFORMATION:
; APPLICANT: Arnold, Andrew
; TITLE OF INVENTION: PRADI Cyclin and its cDNA
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,744
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McConathy, Evelyn H.
; REGISTRATION NUMBER: 35,279
; REFERENCE/DOCKET NUMBER: 0609,4070005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 152 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: not relevant
```

```

; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-460-744-4

Query Match      61.5%; Score 32; DB 3; Length 152;
Best Local Similarity 60.0%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      1 EEVVPXGMSY 10
        ||| | | |
        20 EEVFPPLMNY 29

RESULT 15
US-07-667-711B-4
; Sequence 4, Application US/07667711B
; Patent No. 6110700
; GENERAL INFORMATION:
; APPLICANT: ARNOLD, ANDREW
; TITLE OF INVENTION: Pradi Cyclin and Its cDNA
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVE., NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/667,711B
; FILING DATE: 11-MAR-1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MCPHAIL, DONALD R.
; REGISTRATION NUMBER: 35,811
; REFERENCE/DOCKET NUMBER: 0609,4070000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 152 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; US-07-667-711B-4

Query Match      61.5%; Score 32; DB 3; Length 152;
Best Local Similarity 60.0%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      1 EEVVPXGMSY 10
        ||| | | |
        20 EEVFPPLMNY 29

Search completed: December 22, 2003, 16:43:40
Job time : 20.1333 sec
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:40:14 ; Search time 20.8667 Seconds

(Without alignments)
98.451 Million cell updates/sec

Title: US-09-909-164-5
Perfect score: 52
Sequence: 1 EEVVPXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

```
1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubppaa/US10C_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	73.1	3472	12	US-10-029-120-4
2	38	73.1	3472	12	US-10-027-806-4
3	38	73.1	3472	14	US-10-034-623-4
4	38	73.1	3472	15	US-10-027-801-4
5	36	69.2	153	12	US-10-029-386-32486
6	35	67.3	236	12	US-10-029-386-32076
7	34	65.4	290	12	US-10-094-749-2565
8	34	65.4	947	15	US-10-101-464A-73
9	34	65.4	1163	12	US-10-122-067-4
10	34	65.4	1499	12	US-10-096-534-67
11	34	65.4	1499	12	US-10-122-067-2
12	33	63.5	99	15	US-10-106-698-7477
13	33	63.5	426	15	US-10-214-766-43
14	33	63.5	478	11	US-09-992-600A-108
15	33	63.5	478	11	US-09-924-340-108

16	33	63.5	478	11	US-09-746-783-184	Sequence 184, App
17	33	63.5	478	12	US-09-992-095B-108	Sequence 108, App
18	33	63.5	478	12	US-10-154-678-108	Sequence 108, App
19	33	63.5	478	12	US-09-999-570-108	Sequence 108, App
20	33	63.5	478	15	US-10-000-489-108	Sequence 108, App
21	33	63.5	478	15	US-10-000-986-108	Sequence 108, App
22	33	63.5	648	12	US-10-094-749-3010	Sequence 3010, Ap
23	33	63.5	1152	11	US-09-820-843A-26	Sequence 26, Appl
24	33	63.5	1152	16	US-10-080-170-110	Sequence 130, App
25	33	63.5	1394	12	US-10-116-275-261	Sequence 261, App
26	32	61.5	10	10	US-09-947-387-66	Sequence 66, Appl
27	32	61.5	10	12	US-10-138-375-66	Sequence 66, Appl
28	32	61.5	36	12	US-10-199-820-241	Sequence 241, App
29	32	61.5	242	12	US-10-094-749-2076	Sequence 2076, Ap
30	32	61.5	254	9	US-09-778-927B-53	Sequence 53, Appl
31	32	61.5	276	12	US-10-116-275-114	Sequence 114, App
32	32	61.5	289	10	US-09-919-497-54	Sequence 54, Appl
33	32	61.5	289	14	US-10-024-066-2	Sequence 2, Appl1
34	32	61.5	289	14	US-10-024-066-4	Sequence 4, Appl1
35	32	61.5	292	12	US-10-116-275-275	Sequence 275, App
36	32	61.5	295	10	US-09-925-300-1061	Sequence 1061, Ap
37	32	61.5	529	9	US-09-923-304-4	Sequence 4, Appl1
38	32	61.5	691	9	US-09-925-731-2	Sequence 2, Appl1
39	32	61.5	691	15	US-10-101-921-4	Sequence 10384, A
40	32	61.5	1377	9	US-09-815-242-10384	Sequence 467, App
41	32	61.5	1377	12	US-10-287-274-467	Sequence 4, Appl1
42	32	61.5	2799	14	US-10-151-736-4	Sequence 878, App
43	31	59.6	53	10	US-09-764-847-878	Sequence 878, App
44	31	59.6	53	15	US-10-092-154-878	Sequence 14, Appl
45	31	59.6	59	10	US-09-948-080-14	

ALIGNMENTS

```
RESULT 1
US-10-029-120-4
; Sequence 4, Application US/10029120
; Publication No. US2003015708A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christi
; TITLE OR INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOPI.002A
; CURRENT APPLICATION NUMBER: US/10/029,120
; PRIOR APPLICATION NUMBER: 2001-12-21
; PRIOR FILING DATE: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-029-120-4
```

Query Match 73.1%; Score 38; DB 12; Length 3472;

Best Local Similarity 54.5%; Pred. No. 1.6e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11

Db 2294 EDVVPXGMSYS 2304

RESULT 2
US-10-027-806-4
; Sequence 4, Application US/10027806
; Publication No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.

```
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOB.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-806-4

Query Match      73.1%; Score 38; DB 14; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.6e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMSYS 11
Db      2294 EDVIPRGISFS 2304

RESULT 3
US-10-034-623-4
; Sequence 4, Application US/10034623
; Publication No. US20020198365A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOB.002A
; CURRENT APPLICATION NUMBER: US/10/034,623
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-034-623-4

Query Match      73.1%; Score 38; DB 14; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.6e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMSYS 11
Db      2294 EDVIPRGISFS 2304

RESULT 4
US-10-027-801-4
; Sequence 4, Application US/10027801
; Publication No. US2003005436A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOB.002A
; CURRENT APPLICATION NUMBER: US/10/027,801
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-801-4

Query Match      73.1%; Score 38; DB 15; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.6e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMSYS 11
Db      2294 EDVIPRGISFS 2304

RESULT 5
US-10-029-386-32486
; Sequence 32486, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32486
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF002994.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: SWISSPROT HIT: P43363, EVALU8 5.00e-46
US-10-029-386-32486

Query Match      69.2%; Score 36; DB 12; Length 153;
Best Local Similarity 70.0%; Pred. No. 12;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 EEVVPXGMSY 10
Db      15 KEVPTGHSY 24

RESULT 6
US-10-029-386-32076
; Sequence 32076, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32076
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; OTHER INFORMATION: MAP TO Z98050.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.3
; OTHER INFORMATION: SWISSPROT HIT: P15622, EVALUATE 1.00e-125
US-10-029-386-32076

Query Match      67.3%; Score 35; DB 12; Length 236;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      3 VVPXGMSYS 11
      ||| ||| |||
      80 VVPAGLTYS 88

RESULT 7
US-10-094-749-2565
; Sequence 2565, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TETSUMU
; APPLICANT: OTSUKA, MOTOKUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2565
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2565

Query Match      65.4%; Score 34; DB 12; Length 290;
Best Local Similarity 66.7%; Pred. No. 66;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMS 9
      ||| ||| |||
      35 EEIVPMGIS 43

RESULT 8
US-10-101-464A-73
; Sequence 73, Application US/10101464A
; Publication No. US20030046728A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Strabala, Timothy
; APPLICANT: Mieuwenhuizen, Nicolaas
; APPLICANT: Higgins, Colleen M.
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signaling
; FILE REFERENCE: 11000.1020c2
; CURRENT APPLICATION NUMBER: US/10/101,464A
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 73
; LENGTH: 947
; TYPE: PRT
; ORGANISM: Pinus radiata
US-10-101-464A-73

Query Match      65.4%; Score 34; DB 15; Length 947;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      3 VVPXGMSYS 11
      ||| ||| |||
      686 VMPGGSYS 694

RESULT 9
US-10-122-067-4
; Sequence 4, Application US/10122067
; Publication No. US20030165883A1
; GENERAL INFORMATION:
; APPLICANT: RORY A.J. CURTIS
; TITLE OF INVENTION: 27091, A PHOSPHOLIPID TRANSPORTING
; TITLE OF INVENTION: ATPASE MOLECULE AND USRS THEREFOR
; FILE REFERENCE: MP101-058P1R
; CURRENT APPLICATION NUMBER: US/10/122,067
; CURRENT FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/283434
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1163
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-122-067-4

Query Match      65.4%; Score 34; DB 12; Length 1163;
Best Local Similarity 72.7%; Pred. No. 3.1e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 EEVVPXGMSYS 11
      ||| ||| |||
      133 EEVVPXGMSYS 143

RESULT 10
US-10-096-534-67
; Sequence 67, Application US/10096534
; Publication No. US20030166887A1
; GENERAL INFORMATION:
; APPLICANT: The Brigham and Women's Hospital, Inc.
; APPLICANT: Yates, Karen
; APPLICANT: Mizuno, Shuichi
; APPLICANT: Glowacki, Julie
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF SKELETAL DEGENERATION CONDITIONS
```

```
FILE REFERENCE: B0801/7244/KA/ERP
CURRENT APPLICATION NUMBER: US/10/096,534
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: US 60/274,980
PRIOR FILING DATE: 2001-03-12
NUMBER OF SEQ ID NOS: 79
SOFTWARE: PatentIn version 3.0
SEQ ID NO 67
LENGTH: 1499
TYPE: PRT
ORGANISM: Homo sapiens
US-10-096-534-67

Query Match
Best Local Similarity 65.4%; Score 34; DB 12; Length 1499;
Best Local Similarity 72.7%; Pred. No. 4.2e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
Db 469 EEVVPXGMSYS 479

RESULT 11
US-10-122-067-2
Sequence 2, Application US/10122067
Publication No. US2003016583A1
GENERAL INFORMATION:
APPLICANT: ROY A.J. CURTIS
TITLE OF INVENTION: A PHOSPHOLIPID TRANSPORTING
FILE REFERENCE: MP101-058PIR
CURRENT APPLICATION NUMBER: US/10/122,067
CURRENT FILING DATE: 2002-04-12
PRIOR APPLICATION NUMBER: 60/283434
PRIOR FILING DATE: 2001-04-12
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 1499
TYPE: PRT
ORGANISM: Homo sapiens
US-10-122-067-2

Query Match
Best Local Similarity 65.4%; Score 34; DB 12; Length 1499;
Best Local Similarity 72.7%; Pred. No. 4.2e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
Db 469 EEVVPXGMSYS 479

RESULT 12
US-10-106-698-7477
Sequence 7477, Application US/10106698
Publication No. US20030109690A1
GENERAL INFORMATION:
APPLICANT: RUBEN ET AL.
TITLE OF INVENTION: COLON AND COLON CANCER ASSOCIATED POLYNUCLEOTIDES AND POLYPEPTIDES
FILE REFERENCE: PA005PI
CURRENT APPLICATION NUMBER: US/10/106,698
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 7477
LENGTH: 99
TYPE: PRT
```

```
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (3)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (13)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (14)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (39)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (40)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (57)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (58)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (63)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (65)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (82)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (93)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-7477

Query Match
Best Local Similarity 63.5%; Score 33; DB 15; Length 99;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMSYS 11
Db 73 LVVPGISYS 81

RESULT 13
US-10-214-766-43
Sequence 43, Application US/10214766
Publication No. US20030084473A1
GENERAL INFORMATION:
APPLICANT: GOCAL, GREG
TITLE OF INVENTION: NON-TRANSGENIC HERBICIDE RESISTANT PLANTS
FILE REFERENCE: CA1138
CURRENT APPLICATION NUMBER: US/10/214,766
CURRENT FILING DATE: 2002-08-09
PRIOR APPLICATION NUMBER: US 60/311,734
PRIOR FILING DATE: 2001-08-09
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.2
SEQ ID NO 43
LENGTH: 426
TYPE: PRT
ORGANISM: Vibrio cholerae
US-10-214-766-43

Query Match
Best Local Similarity 63.5%; Score 33; DB 15; Length 426;
Best Local Similarity 60.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSY 10
Db 223 EFVVPAGSY 232
```

RESULT 14

US-09-992-600A-108
; Sequence 108, Application US/09992600A
; Publication No. US20030027161A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US4.DIV
; CURRENT APPLICATION NUMBER: US/09/992,600A
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: JPatent
; SEQ ID NO 108
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-992-600A-108

Query Match

63.5%; Score 33; DB 11; Length 478;
Best Local Similarity 60.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVFXGMSYS 11

Db 239 EVAPAGASTN 248

RESULT 15

US-09-924-340-108
; Sequence 108, Application US/09924340
; Publication No. US20030027248A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US2.REG
; CURRENT APPLICATION NUMBER: US/09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 108
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-924-340-108

Query Match

63.5%; Score 33; DB 11; Length 478;
Best Local Similarity 60.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVFXGMSYS 11

Db 239 EVAPAGASTN 248

Search completed: December 22, 2003, 17:32:40
Job time : 21.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 17:24:36 ; Search time 9.06667 Seconds

(without alignments)
116.675 Million cell updates/sec

Title: US-09-909-164-5

Perfect score: 52

Sequence: 1 EEVVPXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	73.1	3472	2 T31308	hypothetical 367K
2	37	71.2	840	2 T39116	probable sulfate p
3	37	71.2	877	2 T40413	sulfate permease -
4	36	69.2	102	2 A42452	V1 protein - tobac
5	36	69.2	1498	2 B97355	DNA segregation AT
6	35	67.3	225	2 S57810	hypothetical prote
7	35	67.3	425	2 T24111	hypothetical prote
8	35	67.3	670	2 S22293	zinc finger protei
9	35	67.3	749	2 H82691	topoisomerase IV s
10	35	67.3	2717	2 A34203	DNA-binding protei
11	34	65.4	156	2 S54619	hypothetical prote
12	34	65.4	252	2 H69491	cell division inh
13	34	65.4	544	2 C82900	probable ABC subc
14	33	63.5	94	2 I40758	hypothetical prote
15	33	63.5	116	2 B90544	50S ribosomal prot
16	33	63.5	165	2 D94933	hypothetical prote
17	33	63.5	253	2 C81374	hypothetical prote
18	33	63.5	259	2 T34536	hypothetical prote
19	33	63.5	284	2 S75817	hypothetical prote
20	33	63.5	298	2 T47670	beta-ketoacyl-ACP
21	33	63.5	368	2 P72281	hypothetical prote
22	33	63.5	426	2 D82163	3-phosphoshikimate
23	33	63.5	466	2 T43653	cdcd3 protein - fi
24	33	63.5	653	2 D82352	iron(II) ABC tran
25	33	63.5	890	2 A30481	bacteriocin BCNS -
26	33	63.5	1028	2 AF3286	ATP-dependent DNA
27	33	63.5	1152	2 D87046	conserved hypochet
28	33	63.5	1394	2 A35626	transforming growt
29	33	63.5	1401	2 G82336	DNA-directed RNA p

30	33	63.5	1548	2 T04456	hypothetical prote
31	33	63.5	1712	2 A38261	masking protein pr
32	32	61.5	84	2 B97333	hypothetical prote
33	32	61.5	175	2 P00616	transprot protein
34	32	61.5	223	2 T01457	rho protein GDP-di
35	32	61.5	279	2 B72481	hypothetical prote
36	32	61.5	288	2 JC4011	cyclin D2 - rat
37	32	61.5	288	2 I58372	cyclin D2 - rat
38	32	61.5	289	2 A41984	cyclin D2 - mouse
39	32	61.5	289	2 A42822	cyclin D2 - human
40	32	61.5	291	2 S57922	cyclin D2 - Africa
41	32	61.5	291	2 S57925	cyclin D2 - Africa
42	32	61.5	291	2 JC4579	cyclin D1 - chicken
43	32	61.5	291	2 S62730	cyclin D1 - zebra
44	32	61.5	292	2 B42822	cyclin D3 - human
45	32	61.5	295	2 A38977	cyclin D1 - human

ALIGNMENTS

RESULT 1

T31308 hypothetical 367K protein - Cenarchaeum symbiosum

C/Species: Cenarchaeum symbiosum

C/Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000

C/Accession: T31308

R/Schleper, C.; DeLong, E.F.; Preston, C.M.; Feldman, R.A.; Wu, K.Y.; Sanson, R.V.

J. Bacteriol. 180, 5003-5009, 1998

A/Title: Genomic analysis reveals chromosomal variation in natural populations of the

A/Reference number: Z20994; PMID:98422450; PMID:9748430

A/Accession: T31308

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-3472 <SCH>

A/Cross-references: EMBL:AF083072; NID:G3599393; PID:G3599394; PIDN:AAC6269.1

C/Superfamily: Cenarchaeum symbiosum hypothetical 367K protein

Query Match

Best Local Similarity 73.1%; Score 38; DB 2; Length 3472;

Matches 6; Conservativity 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
Db 2294 EDVPRGISFS 2304

RESULT 2

T39116 probable sulfate permease - fission yeast (Schizosaccharomyces pombe)

C/Species: Schizosaccharomyces pombe

C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C/Accession: T39116

R/Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, November 1999

A/Reference number: Z21829

A/Accession: T39116

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-840 <HUN>

A/Cross-references: EMBL:AL132779; PIDN:CAB60015.1; GSPDB:GN00066; SPDB:SPAC869.05C

C/Experimental source: strain 97h-; cosmid c869

A/Genes: SPDB:SPAC869.05C

A/Map position: 1

Query Match 71.2%; Score 37; DB 2; Length 840;

Best Local Similarity 77.8%; Pred. No. 21;

Matches 7; Conservativity 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVVPXGMSYS 11
Db 135 VVVPXGMSYS 143

```

RESULT 3
T40413
sulfate permease - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C/Accession: T40413
R/Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Jimenez Martinez, J.
submitted to the EMBL Data Library, August 1998
A/Reference number: 221926
A/Accession: T40413
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-877 <LN>
A/Cross-references: EMBL:AL031261; PIDN:CAA20298.1; GSPDB:GN00067; SPDB:SPBC3H7.02
A/Experimental source: strain 972h-; cosmid C3H7
C/Genetics:
A/Map position: 2

Query Match
Best Local Similarity 71.2%; Score 37; DB 2; Length 877;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 EVVPXGMSYS 11
DB 148 VVPOGMSYA 156

RESULT 4
A42452
VI protein - tobacco yellow dwarf virus (strain Australia)
C/Species: tobacco yellow dwarf virus
C/Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C/Accession: A42452
R/Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
Virology 187, 633-642, 1992
A/Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow
A/Reference number: A42452; MUID:92188538; PMID:1546458
A/Accession: A42452
A/Molecule type: DNA
A/Residues: 1-102 <MOR>
A/Cross-references: GB:M81103; NID:9335283; PIDN:AAA47947.1; PID:9335284

Query Match
Best Local Similarity 69.2%; Score 36; DB 2; Length 102;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 11
DB 7 QVVPSCINYS 16

RESULT 5
B97355
DNA segregation ATPase, FleK/SpoIIIE family, YVKA B. subtilis ortholog [Imported] - Clost
C/Species: Clostridium acetobutylicum
C/Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C/Accession: B97355
R/Nolling, J.; Bretton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
A/Reference number: A96900; MUID:21359325; PMID:21359325
A/Accession: B97355
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1498 <KUR>
A/Cross-references: GB:AE001437; PIDN:AAK81629.1; PID:G15026814; GSPDB:GN00168
A/Experimental source: Clostridium acetobutylicum ATCC824
C/Genetics:
A/Gene: CAC3709

```

```

Query Match
Best Local Similarity 69.2%; Score 36; DB 2; Length 1498;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
DB 1276 EOKIPWMSY 1285

RESULT 6
S57810
hypothetical protein precursor (clone TPPI1) - tomato
C/Species: Lycopersicon esculentum (tomato)
C/Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
C/Accession: S57810
R/Milligan, S.B.; Gasser, C.S.
Plant Mol. Biol. 28, 691-711, 1995
A/Title: Nature and regulation of pisl1-expressed genes in tomato.
A/Reference number: S57808; MUID:95375233; PMID:7647301
A/Accession: S57810
A/Status: preliminary; nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-225 <WIL>
A/Cross-references: EMBL:U20592; NID:9924625; PIDN:AA80497.1; PID:9924626
C/Superfamily: plant kunitz-type proteinase inhibitor

Query Match
Best Local Similarity 67.3%; Score 35; DB 2; Length 225;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
DB 32 DEVVPNGKTYA 42

RESULT 7
T24111
hypothetical protein R10D12.10 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C/Accession: T24111
R/Percy, C.
submitted to the EMBL Data Library, October 1996
A/Reference number: Z19842
A/Accession: T24111
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-425 <WIL>
A/Cross-references: EMBL:Z81109; PIDN:CAB03241.1; GSPDB:GN00023; CBSP:R10D12.10
A/Experimental source: clone R10D12
C/Genetics:
A/Gene: CBSP:R10D12.10
A/Map position: 5
A/Introns: 23/3; 56/3; 113/3; 257/2

Query Match
Best Local Similarity 67.3%; Score 35; DB 2; Length 425;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
DB 335 EQIVPGLQY 344

RESULT 8
S22293
zinc finger protein AT-BP2 - rat (fragment)
C/Species: Rattus norvegicus (Norway rat)
C/Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 20-Sep-1999
C/Accession: S22293; I78656
R/Mitchellmore, C.; Traboni, C.; Cortese, R.
Nucleic Acids Res. 19, 141-147, 1991

```

A/Title: Isolation of two cDNAs encoding zinc finger proteins which bind to the alpha 1-
A/Reference number: 158280; PMID:9187610; PMID:1901405
A/Accession: S22293
A/Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-670 <MT>
A/Cross-references: EMBL:X54250; NID:957519; PIDN:CAA8151.1; PID:957520
A/Note: the authors did not translate the codon for residue 1
C/Superfamily: HIV-SP2 enhancer-binding protein
C/Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 670;
Best Local Similarity 66.7%; Pred. No. 43;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
DB 376 VVPAGLTYS 384

RESULT 9
H82691
topoisomerase IV subunit XE1353 [imported] - *Xylella fastidiosa* (strain 9a5c)
C/Species: *Xylella fastidiosa*
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
R/Accession: H82691
R/Anonymous: The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequen-
Nature 406, 151-157, 2000
A/Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
A/Reference number: A82515; PMID:2035717; PMID:10910347
A/Note: for a complete list of authors see reference number A59328 below
A/Accession: H82691
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-749 <SM>
A/Cross-references: GB:AE003967; GB:AE003849; NID:9106347; PIDN:AAFP84162.1; GSPDB:GN001
A/Experimental source: strain 9a5c
R/Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvares, R.; B-
Brites, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Carraro, D.M.; Carrier, H-
as-Neto, E.; Docena, C.; El-Dorty, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to Genbank, June 2000
A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm-
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitching, J.P.; Krieger, J.E.; Kurama, E.E.; Laiz-
Chido, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E-
A/Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Mirecca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A-
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Savaas-
A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir-
M.; Tshahko, M.H.; Valleda, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z-
A/Reference number: A59328
A/Contents: annotation
C/Genetics:
A/Map position: 15R
A/Map position: 15R
C/Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain A, phage T4 DNA topoisomerase
A/Accession: XE1353

Query Match 67.3%; Score 35; DB 2; Length 749;
Best Local Similarity 77.8%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
DB 526 EVDPGMSY 534

A/Title: A DNA-binding protein containing two widely separated zinc finger motifs tha-
A/Reference number: A34203; PMID:90169514; PMID:2106471
A/Accession: A34203
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-2717 <FAN>
A/Cross-references: EMBL:X51435; NID:938017; PIDN:CAA35798.1; PID:938018
R/Baldwin Jr., A.S.; LeClair, K.P.; Singh, H.; Sharp, P.A.
Mol. Cell. Biol. 10, 1406-1414, 1990
A/Title: A large protein containing zinc finger domains binds to related sequence ele-
A/Reference number: A34779; PMID:90205817; PMID:2108316
A/Accession: A34779
A/Status: preliminary; nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 801-1072, 'N', 1074-1168, 'K', 1170-1225, 'V', 1227-1434, 'N', 1436-1607, 'T', 1609
A/Cross-references: GB:M32019
C/Superfamily: HIV-SP2 enhancer-binding protein
C/Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 2717;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
DB 2405 VVPAGLTYS 2413

RESULT 11
S54619
hypothetical protein YOR013w - yeast (*Saccharomyces cerevisiae*)
N/Alternate names: hypothetical protein O2612; hypothetical protein YOL303.3
C/Species: *Saccharomyces cerevisiae*
C/Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C/Accession: S54619; S56879
R/de Haan, M.; Maarse, A.C.; Grievell, L.A.
submitted to the EMBL Data Library, May 1995
A/Reference number: S54617
A/Accession: S54619
A/Molecule type: DNA
A/Residues: 1-156 <DEM>
A/Cross-references: EMBL:X87311; NID:91041652; PIDN:CAA60762.1; PID:9829123
R/de Haan, M.; Grievell, L.A.; Maarse, A.C.
submitted to the Protein Sequence Database, July 1996
A/Reference number: S56879
A/Accession: S56879
A/Molecule type: DNA
A/Residues: 1-156 <DEM>
A/Cross-references: EMBL:Z74920; NID:91420109; PIDN:CAA99201.1; PID:91420111; MIPS:YOF-
A/Experimental source: strain S288C
C/Genetics:
A/Cross-references: SCD:S0005539
A/Map position: 15R
C/Superfamily: hypothetical protein YOR013w

Query Match 65.4%; Score 34; DB 2; Length 156;
Best Local Similarity 66.7%; Pred. No. 14;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
DB 50 EVVPLGMDY 58

RESULT 12
H69491
cell division inhibitor (mind-2) homolog - *Archaeoglobus fulgidus*
C/Species: *Archaeoglobus fulgidus*
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 26-Aug-1999
C/Accession: H69491
R/Kien, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dode-
Pleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kitzmann, E.
Glock, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997
 A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes, S. Smith, H.O.; Woose, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
 A:Reference number: A69250; MUID:98049343; PMID:9389475
 A:Accession: H69491
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-252 <KLE>
 A:Cross-references: GB:AE000970; GB:AE000782; NID:g2689293; PIDN:AA89318.1; PID:g264860
 C:Superfamily: cell division inhibitor mind

Query Match 65.4%; Score 34; DB 2; Length 252;
 Best Local Similarity 75.0%; Pred. No. 24;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVFXGMSY 9
 ||| |||
 Db 81 EVIPAGMS 88

RESULT 13
 C82900
 Probable ABC substrate-binding protein, iron UY359 [Imported] - Ureaplasma urealyticum
 C:Species: Ureaplasma urealyticum
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: C82900
 R:Glaes, J.I.; Iefkowitz, E.J.; Glaes, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
 submitted to Genbank, February 2000
 A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min
 A:Reference number: A82870
 A:Accession: C82900
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-544 <GLA>
 A:Cross-references: GB:AE002133; GB:AF222894; NID:g6899339; PIDN:AAF30768.1; GSPDB:GN001
 A:Experimental source: serovar 3; biovar 1
 C:Genetics:
 A:Gene: ABCsbp-5; UY359
 A:Genetic code: SGC3

Query Match 65.4%; Score 34; DB 2; Length 544;
 Best Local Similarity 70.0%; Pred. No. 55;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
 ||| |||
 Db 135 BEVPHYLSY 144

RESULT 14
 140758
 hypothetical protein 1 - Campylobacter jejuni (fragment)
 C:Species: Campylobacter jejuni
 C>Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
 C:Accession: 140758; S47317
 R:Han, E.K.; Chan, V.L.
 J. Bacteriol. 177, 2396-2402, 1995
 A:Title: Expression and characterization of Campylobacter jejuni benzoyl-glycine amidonhy
 A:Reference number: 140758; MUID:95247673; PMID:7730270
 A:Accession: 140758
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-94 <RBS>
 A:Cross-references: EMBL:Z36940; NID:g535805; PIDN:CAA85392.1; PID:g535806

Query Match 63.5%; Score 33; DB 2; Length 94;
 Best Local Similarity 55.6%; Pred. No. 14;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVFXGMSY 10
 ||| |||
 Db 26 DIFPGMSY 34

RESULT 15
 B30544
 50S ribosomal protein L20 [Imported] - Mycoplasma pulmonis (strain UAB CTIP)
 C:Species: Mycoplasma pulmonis
 C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
 C:Accession: B90544
 R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I
 Nucleic Acids Res. 29, 2145-2153, 2001
 A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pu
 A:Reference number: A99512; MUID:21267165; PMID:11353084
 A:Accession: B90544
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-116 <KUM>
 A:Cross-references: GB:AL445566; PID:g14089674; PIDN:CAC13434.1; GSPDB:GN00153
 A:Experimental source: strain UAB CTIP
 C:Genetics:
 A:Gene: MYPu 2610
 A:Genetic code: SGC3
 C:Superfamily: Escherichia coli ribosomal protein L20

Query Match 63.5%; Score 33; DB 2; Length 116;
 Best Local Similarity 77.8%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
 ||| |||
 Db 68 VRPLGMSYS 76

Search completed: December 22, 2003, 17:44:52
 Job time : 11.0667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:43:51 ; Search time 4.6 Seconds

(without alignments)
112.455 Million cell updates/sec

Title: US-09-909-164-5

Perfect score: 52

Sequence: 1 BEVVPXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	38	73.1	1058 1	CARB_FUSNN
2	37	71.2	877 1	SULH_SCHPO
3	36	69.2	102 1	YL1K_TYDVA
4	36	69.2	1498 1	Y1A9_CLOAB
5	35	67.3	2717 1	ZEP1_HUMAN
6	34	65.4	788 1	CY14_NEUCR
7	34	65.4	1499 1	A10C_HUMAN
8	33	63.5	116 1	RL20_MYCPU
9	33	63.5	165 1	Y149_ARCFU
10	33	63.5	253 1	Y990_CAMJE
11	33	63.5	280 1	CTX3_MOUSE
12	33	63.5	466 1	AROA_VIBCH
13	33	63.5	426 1	CC37_SCHPO
14	33	63.5	478 1	GSR2_HUMAN
15	33	63.5	890 1	BCNS_CLOPE
16	33	63.5	1389 1	LTBS_MOUSE
17	33	63.5	1394 1	LTBS_HUMAN
18	33	63.5	1401 1	RPOC_VIBCH
19	33	63.5	1595 1	LTBL_HUMAN
20	33	63.5	1712 1	LTBL_RAT
21	33	63.5	1713 1	LTBL_MOUSE
22	32	61.5	288 1	CGD2_RAT
23	32	61.5	289 1	CGD2_HUMAN
24	32	61.5	289 1	CGD2_MOUSE
25	32	61.5	291 1	CGD1_BRARE
26	32	61.5	291 1	CGD1_XENLA
27	32	61.5	291 1	CGD2_CHICK
28	32	61.5	291 1	CGD2_XENLA
29	32	61.5	292 1	CGD1_CHICK
30	32	61.5	292 1	CGD3_HUMAN
31	32	61.5	295 1	CGD1_HUMAN
32	32	61.5	295 1	CGD1_MOUSE
33	32	61.5	295 1	CGD1_RAT

34	32	61.5	427 1	TOLB_HABIN	P44677 haemophilus
35	32	61.5	529 1	ENP3_HUMAN	O75355 homo sapien
36	32	61.5	691 1	S216_HUMAN	O94616 homo sapien
37	32	61.5	719 1	GSP_CARPA	P90518 cichlida f
38	32	61.5	726 1	PRTP_HSV6U	P52384 human herpe
39	32	61.5	759 1	SCD1_YEAST	P32784 saccharomyc
40	32	61.5	920 1	EDD_RAT	O62671 rattus norv
41	32	61.5	993 1	VLA_TAV	P28931 tomato aspe
42	32	61.5	1377 1	RHSA_ECOLI	P16916 escherichia
43	32	61.5	1397 1	RHSC_ECOLI	P16918 escherichia
44	32	61.5	1411 1	RHSE_ECOLI	P16917 escherichia
45	32	61.5	2799 1	BDP_HUMAN	O95071 homo sapien

ALIGNMENTS

```

RESULT 1
ID CARB_FUSNN STANDARD; PRT; 1058 AA.
AC 08RG86;
AD 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
CN CARB OR FN0422.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteriia; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; Pubmed=11899109;
RA Kapralov V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,
RA Laren N., D'Souza M., Malinas T., Pusch G., Haselkorn R.,
RA Fontein M., Kyrides N., Overbeek R.;
RA "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586."
J. Bacteriol. 184:2005-2016(2002).
CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -1- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -1- PATHWAY: Arginine biosynthesis.
CC -1- PATHWAY: Pyrimidine biosynthesis; first step.
CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL, AB010554; AAL94625.1; ALT_INIT.
DR HAMAP, MF_01210; -; 1.
DR InterPro, IPR006275; CARA_L_glu.
DR InterPro, IPR005483; CPase_L.
DR InterPro, IPR005479; CPase_L_D2.
DR InterPro, IPR005480; CPase_L_D3.
DR InterPro, IPR005481; CPase_L_N.
DR InterPro, IPR004362; MCS_Like.
DR Pfam, PF00289; CPase_L_chain; 2.
DR Pfam, PF02786; CPase_L_D2; 2.
DR Pfam, PF02787; CPase_L_D3; 1.

```

DR Pfam; PF02142; MCS; 1.
 DR PRINTS; PR00098; CPSASE.
 DR TIGRPFAMS; TIGR01369; CPSASE_1; 1.
 DR PROSITE; PS00866; CPSASE_1; 2.
 DR PROSITE; PS00867; CPSASE_2; 2.
 KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 KM ATP-binding; Manganese; Complete proteome.
 FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 2 546 OLGOMERIZATION DOMAIN.
 FT DOMAIN 3 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 4 1058 ALLOSTERIC DOMAIN.
 FT REPEAT 1 546
 FT REPEAT 2 1058
 FT NP_BIND 153 1058
 FT NP_BIND 302 352
 FT METAL 284 284
 FT METAL 298 298
 FT METAL 300 300
 FT METAL 820 820
 FT METAL 832 832
 SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1E39F CRC64;

Query Match 73.1%; Score 38; DB 1; Length 1058;
 Best Local Similarity 60.0%; Pred. No. 6.3;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPMGMSYS 11
 DB 190 EIVPGLNLS 199

RESULT 2

ID SUH_SCHPO STANDARD; PRT; 877 AA.
 AC 074377;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable sulfate permease CH7.02.
 GN SPBC3H7.02.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 RN NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Baeham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean C.,
 RA Mooney P., Mouton S., Mungall K., Murphy L., Niblett D., Odeli C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitz E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Gymnopoulos B.,
 RA Welteens I., Vansteegle E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Filiz C., Holtzer E., Moestl D., Hiltbert H.,
 RA Borzym K., Langer I., Beck A., Leirich H., Reinhardt R., Pohl T.M.,
 RA Rager P., Zimmermann W., Wedler H., Wambut R., Purnelle B.,
 RA Goffeau A., Cailliau E., Dreame S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Ayes S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.;

RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 CC -1- FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE INTO THE CELL (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
 CC -1- SIMILARITY: BELONGS TO THE SLC26A FAMILY OF TRANSPORTERS.
 CC -1- SIMILARITY: Contains 1 STAS domain.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).

DR EMBL; AL031261; CAA20298.1; -.
 DR PIR; T40413; T40413.
 DR GenDB; SPombe; SPBC3H7.02; -.
 DR InterPro; IPR002645; STAS.
 DR InterPro; IPR01902; Sulph_transpc.
 DR Pfam; PF01740; STAS; 1.
 DR TIGRPFAMS; TIGR00815; sulP; 1.
 DR PROSITE; PS01130; SLC26A; 1.
 DR PROSITE; PS00801; STAS; 1.

KW Transport; Transmembrane.
 FT TRANSMEM 133 153 POTENTIAL.
 FT TRANSMEM 161 181 POTENTIAL.
 FT TRANSMEM 186 206 POTENTIAL.
 FT TRANSMEM 221 241 POTENTIAL.
 FT TRANSMEM 243 263 POTENTIAL.
 FT TRANSMEM 292 312 POTENTIAL.
 FT TRANSMEM 329 349 POTENTIAL.
 FT TRANSMEM 384 404 POTENTIAL.
 FT TRANSMEM 424 444 POTENTIAL.
 FT TRANSMEM 461 481 POTENTIAL.
 FT TRANSMEM 484 504 POTENTIAL.
 FT TRANSMEM 518 538 POTENTIAL.
 FT TRANSMEM 543 563 POTENTIAL.
 FT DOMAIN 594 747 STAS.
 SQ SEQUENCE 877 AA; 96373 MW; 56995A8493371E43 CRC64;

Query Match 71.2%; Score 37; DB 1; Length 877;
 Best Local Similarity 77.8%; Pred. No. 8.4;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPMGMSYS 11
 DB 148 VVPMGMSYA 156

RESULT 3

ID Y1LK TYDVA STANDARD; PRT; 102 AA.
 AC P31619;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-1993 (Rel. 27, Last annotation update)
 DE Hypothetical 11.2 kDa protein.
 GN Y1.
 OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
 OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
 RN NCBI_TaxID=31599;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92188538; PubMed=1546458;
 RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.B.;
 RT "The nucleotide sequence of the infectious cloned DNA component of
 RT tobacco yellow dwarf virus reveals features of geminiviruses
 RT infecting monocotyledonous plants.";
 RL Virology 187:633-642(1992).
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M81103; AAA47947.1; -
 DR PIR; A42452; A42452.
 DR InterPro: IPR002621; Gemini_mov.
 DR Pfam; PF01708; Gemini_mov; I.
 DR Hypothetical protein.
 KW SEQUENCE 102 AA; 11178 MW; A40BCF1E0AF5B67 CRC64;

Query Match 69.2%; Score 36; DB 1; Length 102;
 Best Local Similarity 60.0%; Pred. No. 1.4;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPGXGMSYS 11
 DB 7 QVVPSCINYS 16

RESULT 4
 ID Y1A9_CLOAB STANDARD; PRT; 1498 AA.
 AC 004351;
 DT 01-PEB-1994 (Rel. 28, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hypothetical protein CAC3709.
 GN CAC3709
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21359325; PubMed=1146286;
 RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum.";
 RL J. Bacteriol. 183:4823-4838(2001).
 RN [2]
 RP SEQUENCE OF 1-108 FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=93273706; PubMed=8501044;
 RA Sauer U., Duerre P.;
 RT "Sequence and molecular characterization of a DNA region encoding a
 RT small heat shock protein of Clostridium acetobutylicum.";
 RL J. Bacteriol. 175:3394-3400(1993).
 CC [1]- SIMILARITY: Contains 2 FtsK/SpoIIIE domains.
 CC [1]- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
 CC in positions 76 and 106.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR PROSITE; P850901; FTSK_SPOIIIE; 2.
 KW Hypothetical protein; ATP-binding; Complete proteome; Repeat.
 FT DOMAIN 655 857 FTSK/SPOIIIE 1.
 FT NP BIND 675 682 ATP (POTENTIAL).
 FT DOMAIN 1001 1188 FTSK/SPOIIIE 2.
 SQ SEQUENCE 1498 AA; 168968 MW; FF42037A335A9649 CRC64;

Query Match 69.2%; Score 36; DB 1; Length 1498;
 Best Local Similarity 60.0%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVDPXGMSY 10
 DB 1276 EQKIPMGMSY 1285

RESULT 5
 ID ZEP1_HUMAN STANDARD; PRT; 2717 AA.
 AC P15832;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein 40 (Human immunodeficiency virus type I enhancer-
 DE binding protein 1) (HIV-Ep1) (Major histocompatibility complex binding
 DE protein 1) (MBP-1) (Positive regulatory domain II binding factor 1)
 DE (PRDIT-BF1).
 GN HIVEP1 OR ZNF40.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90169514; PubMed=2106471;
 RA Fan C.M., Maniatis T.;
 RT "A DNA-binding protein containing two widely separated zinc finger
 RT motifs that recognize the same DNA sequence.";
 RL Genes Dev. 4:29-42(1990).
 RN [2]
 RP STRUCTURE BY NMR OF 2113-2142.
 RX MEDLINE=91064333; PubMed=2248949;
 RA Omichinski J.G., Clore G.M., Appella E., Sakaguchi K.,
 RA Gronenborn A.M.;
 RT "High-resolution three-dimensional structure of a single zinc finger
 RT from a human enhancer binding protein in solution.";
 RL Biochemistry 29:9324-9334(1990).
 RN [3]
 RP STRUCTURE BY NMR OF 2087-2142.
 RX MEDLINE=92232684; PubMed=1567844;
 RA Omichinski J.G., Clore G.M., Roblen M., Sakaguchi K., Appella E.,
 RA Gronenborn A.M.;
 RT "High-resolution solution structure of the double Cys2His2 zinc
 RT finger from the human enhancer binding protein MBP-1.";
 RL Biochemistry 31:3907-3917(1992).
 CC [1]- FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE
 CC 5'-GGGAGCTTC-3' WHICH IS FOUND IN THE ENHANCER ELEMENTS OF
 CC NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1.
 CC IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS
 CC OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I
 CC MHC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT
 CC IN T-CELL ACTIVATION.
 CC [1]- SUBCELLULAR LOCATION: Nuclear.
 CC [1]- INDUCTION: BY MITOGEN AND PHORBOL ESTER.
 CC [1]- DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY
 CC SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH
 CC ZINC-FINGER IN-BETWEEN.
 CC [1]- SIMILARITY: STRONG, TO HIVEP2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X51435; CAA35798.1; -.
DR PIR; A34203; A34203.
DR PDB; 3ZNF; 15-JAN-92.
DR PDB; 4ZNF; 15-JAN-92.
DR PDB; 1BRO; 31-OCT-93.
DR TRANSFAC; T00497; -.
DR Genem; HGNC:4920; HIVEP1.
DR MIM; 194540; -.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0003677; R:DNA binding activity; TAS.
DR InterPro; IPR007087; ZNF_C2H2.
DR Pfam; PF00096; ZF_C2H2; 5.
DR SMART; SM00355; ZNF_C2H2; 4.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
KM Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KM Nuclear protein; Repeat; 3d-structure.
FT ZN_FING 406 428 C2H2-TYPE.
FT ZN_FING 434 456 C2H2-TYPE.
FT ZN_FING 958 981 C2H2-TYPE. (POTENTIAL).
FT ZN_FING 2087 2109 C2H2-TYPE.
FT ZN_FING 2115 2139 C2H2-TYPE.
FT DOMAIN 803 806 POLY-SER.
FT STRAND 2088 2088
FT TURN 2090 2092
FT STRAND 2095 2095
FT HELIX 2099 2108
FT TURN 2109 2109
FT STRAND 2115 2116
FT STRAND 2123 2124
FT HELIX 2127 2135
SQ SEQUENCE 2717 AA; 297217 MW; D45D3CA951FEA561 CRC64;

Query Match 67.3%; Score 35; DB 1; Length 2717;
Best Local Similarity 66.7%; Pred. No. 73;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
Db 2405 VVPAGLTVS 2413

RESULT 6
CY14_NEUCR STANDARD; PRT; 788 AA.
AC P23622;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sulfate permease II.
GN CYS-14.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91129256; PubMed=1825178;
RA Ketter J.S., Jatal G., Fu Y.-H., Marzluft G.A.;
RT "Nucleotide sequence, messenger RNA stability, and DNA recognition
RT elements of cys-14, the structural gene for sulfate permease II in
RT Neurospora crassa.";
RL Biochemistry 30:1780-1787(1991).
RN [2]
RP PROBABLE REVISIONS.
RX MEDLINE=94188926; PubMed=8140616;
RA Sandal N.N., Marcker K.A.;
RT "Similarities between a soybean nodulin, Neurospora crassa sulphate
RT permease II and a putative human tumour suppressor.";
```

```
RL Trends Biochem. Sci. 19:19-19(1994).
CC - FUNCTION: UPTAKE OF SULFATE INTO THE CELL.
CC - SUBCELLULAR LOCATION: Integral membrane protein.
CC - INDUCTION: Highly expressed, but only in cells subject to sulfur
CC limitation, and it is turned on by the positive-acting Cys-3
CC sulfur regulatory protein.
CC - MISCELLANEOUS: SULFATE PERMEASE II IS MAINLY FOUND IN MYCELSIA.
CC - SIMILARITY: BELONGS TO THE SLC26A FAMILY OF TRANSPORTERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M59167; AAA3615.1; ALT SEQ.
DR InterPro; IPR001902; Sulfate transp.
DR Pfam; PF00916; Sulfate_transp; 1.
DR TIGRfams; TIGR00815; sulP; 1.
DR PROSITE; PS01130; SLC26A; 1.
KM Transport; Transmembrane; Glycoprotein.
FT TRANSMEM 71 91 POTENTIAL.
FT TRANSMEM 103 123 POTENTIAL.
FT TRANSMEM 128 148 POTENTIAL.
FT TRANSMEM 171 191 POTENTIAL.
FT TRANSMEM 193 213 POTENTIAL.
FT TRANSMEM 271 291 POTENTIAL.
FT TRANSMEM 326 346 POTENTIAL.
FT TRANSMEM 363 383 POTENTIAL.
FT TRANSMEM 451 471 POTENTIAL.
FT TRANSMEM 474 494 POTENTIAL.
FT CARBOHYD 23 23
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 788 AA; 87864 MW; 4FC04B60798CE77 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 788;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
Db 90 VVPQGMAYR 98

RESULT 7
A10C_HUMAN STANDARD; PRT; 1499 AA.
AC O60312; Q96914;
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Potential phospholipid-transporting ATPase VC (EC 3.6.3.1) (ATPVC)
DE (Aminophospholipid translocase VC).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21225279; PubMed=11326269;
RA Oshimura M., Kaishiwagi A., Mitsuoka K., Nakao M., Kondo I., Saitoh S.,
RA Oshimura M.;
RT "A novel maternally expressed gene, ATP10C, encodes a putative
RT aminophospholipid translocase associated with Angelman syndrome.";
RL Nat. Genet. 28:19-20(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=2131119; PubMed=11353404;
RA Herzog L.B.K., Kim S.-U., Cook E.H. Jr., Ledbetter D.H.;
RT "The human aminophospholipid-transporting ATPase gene ATP10C maps
```

RT adjacent to UB32A and exhibits similar imprinted expression.";
 RL Am. J. Hum. Genet. 68:1501-1505(2001).
 RN [3]
 RP SEQUENCE OF 337-1499 FROM N.A.
 RC TISSUE=Brain;
 RA MEDLINE=98290545; PubMed=9628581;
 RX Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
 RA Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human gene. IX.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 5:31-39(1998).
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -1- TISSUE SPECIFICITY: Wide expression, with highest levels in
 CC kidney, followed by lung, brain, prostate, testis, ovary, and
 CC small intestine.
 CC -1- DISEASE: Defects in ATP10C are a cause of Angelman syndrome (AS)
 CC [MIM:105830]; also known as 'happy puppet syndrome'. It is
 CC characterized by features of severe motor and intellectual
 CC retardation, microcephaly, ataxia, frequent jerky limb movements
 CC and flapping of the arms and hands, hypotonia, hyperactivity,
 CC hypopigmentation, seizures, absence of speech, frequent smiling
 CC and episodes of paroxysmal laughter, and an unusual facies
 CC characterized by macrostomia, a large mandible and open-mouthed
 CC expression, a great propensity for protruding the tongue ('tongue
 CC thrusting'), and an occipital groove.
 CC -1- SIMILARITY: Belongs to the cation transport ATPases family (P-type
 CC ATPases). Subfamily IV.
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB051358; BAB47392.1; -;
 DR EMBL; AY029504; AAK33100.1; JOINED.
 DR EMBL; AY029487; AAK33100.1; JOINED.
 DR EMBL; AY029488; AAK33100.1; JOINED.
 DR EMBL; AY029489; AAK33100.1; JOINED.
 DR EMBL; AY029490; AAK33100.1; JOINED.
 DR EMBL; AY029491; AAK33100.1; JOINED.
 DR EMBL; AY029492; AAK33100.1; JOINED.
 DR EMBL; AY029493; AAK33100.1; JOINED.
 DR EMBL; AY029494; AAK33100.1; JOINED.
 DR EMBL; AY029495; AAK33100.1; JOINED.
 DR EMBL; AY029496; AAK33100.1; JOINED.
 DR EMBL; AY029497; AAK33100.1; JOINED.
 DR EMBL; AY029498; AAK33100.1; JOINED.
 DR EMBL; AY029499; AAK33100.1; JOINED.
 DR EMBL; AY029500; AAK33100.1; JOINED.
 DR EMBL; AY029501; AAK33100.1; JOINED.
 DR EMBL; AY029502; AAK33100.1; JOINED.
 DR EMBL; AY029503; AAK33100.1; JOINED.
 DR EMBL; AB011138; BAA25492.1; -;
 DR Genew; HGNC:13547; ATP10C.
 DR MIM; 605855; -;
 DR MIM; 105830; -;
 DR GO; GO:0016071; C:Integral to membrane; NAS.
 DR GO; GO:0004012; F:Phospholipid-translocating ATPase activity; NAS.
 DR GO; GO:0008360; P:Regulation of cell shape; NAS.
 DR InterPro; IPR001757; ATPase_EI-R2.
 DR InterPro; IPR006539; F1IIPase.
 DR InterPro; IPR005834; Hydrolase.
 DR Pfam; PF00702; Hydrolase_1.
 DR PRINTS; PR00119; CATATPASE.
 DR TIGRFAMs; TIGR01652; ATPase_Plipid_1.
 DR TIGRFAMs; TIGR01494; ATPase_P-type_6.
 DR PROSITE; PS00154; ATPase_EI_R2; 1.
 KW Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;

KW Multigene family.
 FT DOMAIN 1 86
 FT TRANSMEM 87 106
 FT DOMAIN 107 110
 FT TRANSMEM 111 128
 FT DOMAIN 129 309
 FT TRANSMEM 310 332
 FT DOMAIN 333 362
 FT TRANSMEM 363 384
 FT DOMAIN 385 1087
 FT TRANSMEM 1088 1108
 FT DOMAIN 1109 1119
 FT TRANSMEM 1120 1140
 FT DOMAIN 1141 1170
 FT TRANSMEM 1171 1192
 FT DOMAIN 1193 1199
 FT TRANSMEM 1200 1222
 FT DOMAIN 1223 1228
 FT TRANSMEM 1229 1249
 FT DOMAIN 1250 1267
 FT TRANSMEM 1268 1292
 FT DOMAIN 1293 1499
 FT MOD_RES 427 427
 FT METAL 1031 1031
 FT METAL 1035 1035
 FT DOMAIN 467 470
 FT CONFLICT 388 388
 FT SEQUENCE 1499 AA; 167687 MW; DA996A4D0635A68D CRC64;
 QY 1 EEVVPKGMYS 11
 Db 469 EEVVPKGMYS 479
 Query Match 65.4%; Score 34; DB 1; Length 1499;
 Best Local Similarity 72.7%; Pred. No. 64;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 RESULT 8
 RL20 MYCPU STANDARD; PRT; 116 AA.
 AC Q980V0;
 DT 26-FEB-2003 (Rel. 41, Created)
 DT 26-FEB-2003 (Rel. 41, Last sequence update)
 DT 26-FEB-2003 (Rel. 41, Last annotation update)
 DE 50S ribosomal protein L20.
 GN RPLT OR MYPV 2610.
 OS Mycoplasma pulmonis.
 CC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2107;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UAB CTIP;
 RX MEDLINE=21267165; PubMed=11353084;
 RA Chambaud I., Hellig R., Ferris S., Barbe V., Samson D., Galisson F.,
 RA Moszer I., Dybvig K., Wroblewski H., Viart A., Kocha E.P.C.,
 RA Blanchard A.;
 RT "The complete genome sequence of the murine respiratory pathogen
 RT Mycoplasma pulmonis.";
 RL Nucleic Acids Res. 29:2145-2153(2001).
 CC -1- FUNCTION: This protein binds directly to 23S ribosomal RNA and is
 CC necessary for the in vitro assembly process of the 50S ribosomal
 CC subunit. It is not involved in the protein synthesizing functions
 CC of that subunit (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE L20P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

CC -----
DR EMBL: AL445563; CAC13434.1; -.
DR PIR: E90544; E90544.
DR MyPublic; MYPU_2610; -.
DR HAMAP: MF_00382; -. 1.
DR InterPro: IPR005813; L20.
DR InterPro: IPR005812; L20_bact.org.
DR Pfam: PF00453; Ribosomal_L20; 1.
DR PRINTS: PR00062; RIBOSOMALL20.
DR ProDom: PD002389; L20; 1.
DR TIGRPFAM: TIGR01032; rplT_bact; 1.
DR POSITIVE: PS00937; RIBOSOML_L20; 1.
KM Ribosomal protein, rRNA-binding, Complete proteome.
SQ SEQUENCE 116 AA; 13565 MW; C59C748901B18F14 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 116;
Best Local Similarity 77.8%; Pred. No. 7.1;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 EVVPGMSYS 11
DB 68 VRPLGMSYS 76

RESULT 9
ID YJ49_ARCFU STANDARD; PRT; 165 AA.
AC 028330;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Hypothetical protein AF1949.
GN AF1949.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RC MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.B.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervagge A.R., Graham D.E., Kyrtides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirsnes E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 350:364-370(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000968; AAB89307.1; -.
DR PIR: D69493; D69493.
DR TIGR: A1949; -.
KM Hypothetical protein, Transmembrane, Complete proteome.
FT TRANSMEM 7 27 POTENTIAL.
SQ SEQUENCE 141 161
SQ SEQUENCE 165 AA; 17588 MW; BBCL7054810ADB8 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 165;

```

```

Best Local Similarity 60.0%; Pred. No. 10;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
DB 60 EESIPDGASY 69

RESULT 10
ID Y990_CAMJF STANDARD; PRT; 253 AA.
AC P454B9; O9PNU0;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein Cj0990C.
GN Cj0990C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NOTC 11168;
RC MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltham D., Holt R.D.,
RA Jorgensen K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrett B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
RN [2]
RP SEQUENCE OF 160-253 FROM N.A.
RC STRAIN=ATCC 43431 / TGH 9011;
RC MEDLINE=95247673; PubMed=7730270;
RA Hani B.K., Chan V.L.;
RT "Expression and characterization of Campylobacter jejuni
RT benzoylglutamate amidohydrolase (Hippuricase) gene in Escherichia
RT coli."
RL J. Bacteriol. 177:2396-2402(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL139076; CAB73246.1; -.
DR EMBL: Z36940; CAAB5392.1; -.
DR PIR: C81374; C81374.
DR PIR: I40758; I40758.
KM Hypothetical protein, Complete proteome.
SQ SEQUENCE 253 AA; 29783 MW; F96D3FF3255F8A6A CRC64;

Query Match 63.5%; Score 33; DB 1; Length 253;
Best Local Similarity 55.6%; Pred. No. 16;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EEVVPXGMSY 10
DB 185 DIFPGMSY 193

RESULT 11
ID CTX3_MOUSE STANDARD; PRT; 280 AA.
AC Q9D387; Q9CXQ4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)

```

15-SEP-2003 (Rel. 42, Last annotation update)
 DE Protein C20orf103 homolog precursor.
 GN C20ORF103.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
 RX MEDLINE=2108560; PubMed=11217851;
 RA Kawai J., Shitagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Nishitani Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,
 RA Saito T., Okazaki Y., Gajodori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
 RA Kiehl P., Lewis S., Matsuo Y., Nikaido I., Pebole G., Quackenbush J.,
 RA Schirrl L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Kono H., Baldarelli R., Baren G.,
 RA Blake J., Botfeill D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohzuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RP TISSUE=Mammary fibroblast;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klauener R.D., Collins F.S., Wagner L., Shenman C.M., Schuler G.D.,
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat H.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
 RA Spatton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Bha S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.U., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Hellon E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Scheraga A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9DJ87-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9DJ87-2; Sequence=VSP_003820;
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
 CC in positions 174 and 239.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

DR EMBL; AK014127; BAB29169.1; -;
 DR EMBL; AK018322; BAB31124.1; ALT_FRAME.
 DR EMBL; BC004791; AA004791.1;
 DR MGD; MGI:1920366; 3110035N03RIK.
 DR MGD; MGI:1923411; 6330527006RIK.
 KW Transmembrane; Signal; Alternative splicing.
 FT SIGNAL 1 29
 FT CHAIN 30 280
 FT DOMAIN 30 235
 FT TRANSLEM 236 256
 FT DOMAIN 257 280
 FT CARBOHYD 35 35
 FT CARBOHYD 53 53
 FT CARBOHYD 102 102
 FT CARBOHYD 127 127
 FT VARSPLIC 1 118
 FT CONFLICT 221 221
 FT CONFLICT 230 230
 FT CONFLICT 238 238
 SQ SEQUENCE 280 AA; 31721 MW; FA1ID7BP9F5CCBP CRC64;
 Query Match 63.5%; Score 33; DB 1; Length 280;
 Best Local Similarity 75.0%; Pred. No. 18;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 3 VDPXGMSY 10
 Db 173 VTPAGMSY 180
 RESULT 12
 AROA_VIBCH STANDARD; PRT; 426 AA.
 ID AROA_VIBCH
 AC Q9KRB0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
 DE enolpyruvylshikimate-3-phosphate synthase) (EPSPS) (EPSPS).
 GN AROA OR VC1732.
 OS Vibrrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrrio.
 NC NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=El Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Haft D.H., Hickley B.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA Salzberg S.L., Ustebach T., Fleischmann R.D., Nieman W.C., White O.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
 RT cholerae.";
 RL Nature 406:477-483(2000).
 CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
 CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
 CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
 CC sixth step.
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- SIMILARITY: Belongs to the EPSP synthase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

CC or send an email to license@isb-sib.ch).
CC EMBL; AE004251; AAF94882.1; -.
DR PIR; D82163; D82163.
DR TIGR; VC1732; -.
DR HAMAP; MF_00210; -.
DR InterPro; IPR006264; Archa.
DR InterPro; IPR001986; BPS_synthase.
DR Pfam; PF00275; BPS_synthase.
DR Prodom; PD001867; BPS_synthase; 1.
DR TIGRfam; TIGR01356; arca; 1.
DR PROSITE; PS00104; BPS_SYNTHASE_1; 1.
DR PROSITE; PS00885; BPS_SYNTHASE_2; 1.
KW Aromatic amino acid biosynthesis; Transferase; Complete proteome.
SQ SEQUENCE 426 AA; 46101 MM; 38852D64835F81C3 CRC64;

Query Match
Best Local Similarity 63.5%; Score 33; DB 1; Length 426;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DY 1 EEVVPXGMSY 10
DB 223 EFVIFAGQSY 232

RESULT 13
CC37_SCHPO STANDARD; PRT; 466 AA.
ID CC37_SCHPO
AC 094740;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hsp90 co-chaperone Cdc37 (Hsp90 chaperone protein kinase-targeting subunit) (Cell division control protein 37).
GN CDC37 OR SPBC9B6.10.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; Schizosaccharomycetaceae; Schizosaccharomycetes.
OC NCB1_TaxID=4896;
RX MEDLINE=21648401; PubMed=11859360;
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A., Sgouras J., Peat N., Hayles J., Baker S., Baaham D., Bowman S., Brooks M., Brown D., Brown S., Chillingworth T., Churcher C.M., Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A., Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson A., Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K., James K., Jones L., Jones M., Leather S., McDonald S., McLean J., Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C., Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E., Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S., Skelton J., Simmonds M., Squares R., Squares S., Stevens K., Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S., Woodward J., Volkart G., Aert R., Robben J., Gymnopoulos B., Wellens I., Vansteelt E., Rieger M., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fritz C., Holzer E., Moser D., Hilbert H., Bazer M., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M., Beer P., Zimmermann W., Medler H., Wambutt R., Purnelle B., Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S., Calbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., Lucas R.R., Rochet W., Galliard C., Tallada V.A., Garzon A., Thode G., Daga R.R., Chuado L., Jimenez J., Sanchez M., del Rey F., Benito J., Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L., Carnitiz L., Lowe T., McCombie W.R., Paulsen I., Potashkin J., Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.,

```

```

RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -1- FUNCTION: With Hsp90 it forms a complex that binds to several kinases, resulting in stabilization and promotion of their activity (By similarity).
CC -1- SUBUNIT: Forms a complex with Hsp90. Interacts with a number of kinases (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE CDC37 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL; AJ132377; CAB38758.1; -.
DR EMBL; AJ132376; CAB38757.1; -.
DR EMBL; AL049763; CAB42371.2; -.
DR PIR; T43653; T43653.
DR GeneDB; SPombe; SPBC9B6.10; -.
DR InterPro; IPR004918; Cdc37.
DR Pfam; PF03234; Cdc37; 1.
KW Chaperone; Cell division; Cell cycle.
SQ SEQUENCE 466 AA; 52554 MM; 64723B34CAB3C5 CRC64;

Query Match
Best Local Similarity 63.5%; Score 33; DB 1; Length 466;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

DY 1 EEVVPXGMSY 10
DB 98 DSAIPGMSY 107

RESULT 14
ID GSR2_HUMAN STANDARD; PRT; 478 AA.
AC GSR2_HUMAN
GN GSR2; Q9BTC6; Q9HAX6; Q9NPP1; Q9NPP4; Q9UP12;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glioma tumor suppressor candidate region gene 2 protein (p60).
GN GLTSCR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCB1_TaxID=9606;
RX MEDLINE=20175430; PubMed=10708517;
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stappleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Schetz T.E., Brownstein M.D., Ustin T.B., Toshiyuki S., Carnitiz P., Prange C., Bosa S.S., Locquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Roha S.A., McGowan P.J., McKernan K.J., Malek J.A., Gnatatne P.H., Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hultj S.W.,

```

RA	Villation D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs S.A.,
RA	Fahney J., Helton E., Kettelman M., Madan A., Rodrigues R.,
RA	Whiting M., Madan A., Young A.C., Shervenchenko Y., Bouffard G.G.,
RA	Bikesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA	Rodriguez A.C., Grimwood J., Schmitz J., Myers R.M.,
RA	Butterfield Y.S.N., Krzyzinski M.I., Skalske U., Smallus D.E.,
RA	Schnecker A., Schain J.E., Jones S.J.M., Marra M.A.,
RT	"Generation and initial analysis of more than 15,000 full-length
RT	human and mouse cDNA sequences,"
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN	[3]
RP	SEQUENCE OF 9-478 FROM N.A., AND SUBCELLULAR LOCATION.
RX	MEDLINE=99214318; PubMed=10196275;
RA	Bruni R., Fineschi B., Ogle W.O., Roizman B.;
RT	"A novel cellular protein, p60, interacting with both herpes simplex
RT	virus 1 regulatory proteins ICP22 and ICP0 is modified in a
RT	cell-type-specific manner and is recruited to the nucleus after
RT	infection."
RL	J. Virol. 73:3810-3817(1999).
RN	[4]
RP	SEQUENCE OF 12-478 FROM N.A.
RA	Andreu N., Estivill X., Escarceller M., Sunmy L.;
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN	[5]
RP	SEQUENCE OF 218-477 FROM N.A.
RA	TISSUE-Testis;
RC	Pousterka A., Klein M., Newes H.-W., Gassenhuber J., Wiemann S.;
RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC	-1- SUBUNIT: Interacts with HSV-1 early proteins ICP27 and ICP0.
CC	-1- SUBCELLULAR LOCATION: Nuclear.
CC	-1- TISSUE SPECIFICITY: Expressed at high levels in heart and
CC	pancreas, moderate levels in placenta, liver, skeletal muscle, and
CC	kidney, and low levels in brain and lung.
CC	-1- SIMILARITY: BELONGS TO THE GLTSCR2 FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC	or send an email to license@sib-sib.ch).
CC	-----
DR	EMBL; AF182076; AAFC2873.1; -
DR	EMBL; BC004229; AAH04229.1; -
DR	EMBL; BC006311; AAH06311.1; -
DR	EMBL; BC010095; AAAH0095.1; -
DR	EMBL; AF296124; AAC30413.1; -
DR	EMBL; AL359335; CAB94786.1; -
DR	EMBL; AL359336; CAB94787.1; -
DR	EMBL; AL122063; CAB59242.1; -
DR	SWISS-2DPAGE; Q9NZM5; HUMAN.
DR	GeneW; HGNC:4333; GLTSCR2.
DR	MM, 605651;
DR	GO; GO:0008181; P:tumor suppressor; TAS.
KW	Nuclear Protein; Polymorphism.
FT	VARIANT 389 389
FT	R -> O.
FT	/FTID=VAR_011486.
FT	GGS -> HEQ (IN REF. 2; AAH04229).
FT	D -> H (IN REF. 3)
FT	RRKEQLRPFKSGFORNMTEPRERAKFRKKYKVULVEKRAF
FT	PKGDNPAQRITA (IN REF. 3).
FT	SDNPDLRYLVGDDEFFLE -> LNNDPDKRVVPMGLTFPG
FT	(IN REF. 3).
FT	A -> S (IN REF. 2; AAH04229).
FT	D -> H (IN REF. 3)
FT	PEGNLIRDFKSGFORNMTEPRERAKFRKKYKVULVEKRAF
FT	REIQ -> VLTVSCKGACFPWTSLFVPFGGRHHGCP
FT	WAGVGPMDRG (IN REF. 5).
FT	EGNILRDFKSGFORNMTEPRERAKFRKKYKVULVEKAFR
FT	ETOL -> RGHSEFTSGRARCGI (IN REF. 3).
FT	7FI89Z3BE348CB52B CXC64;
SO	SEQUENCE 478 AA; 54417 MW;

Query Match	Similarity	Score	DB	Length
Best Local	60.0%	Pred. No. 31		
Matches	6	Conservative	1	Mismatches
			3	Indels
			0	Gaps

Qy	2	EVEXPGENSYS	11
	239	EVAPAGASYN	248

Db	239	EVAPAGASYN	248

RESULT	15
BCNS_CLOPE	STANDARD;
ID	BCNS_CLOPE
AC	P08696;
DT	01-JAN-1988 (Rel. 06, Created)
DT	01-JAN-1988 (Rel. 06, Last sequence update)
DT	01-OCT-1994 (Rel. 30, Last annotation update)
DE	Bacteriocin BCNS.
GN	Clostridium perfringens.
OS	Plasmid pip404.
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC	Clostridium.
OX	NCBI_TaxID=1502;
XX	[1]
RP	SEQUENCE FROM N.A.
RP	STRAIN=CPN50;
RC	MEDLINE=88336297; PubMed=2901768;
FX	Garnier T., Cole S.T.;
RA	"Complete nucleotide sequence and genetic organization of the bacteriocinogenic plasmid, pip404, from Clostridium perfringens.";
RT	Plasmid 19:134-150(1988).
RL	[2]
RN	SEQUENCE FROM N.A.
RP	STRAIN=CPN50;
RC	MEDLINE=87057020; PubMed=2877971;
FX	Garnier T., Cole S.T.;
RA	"Characterization of a bacteriocinogenic plasmid from Clostridium perfringens and molecular genetic analysis of the bacteriocin-encoding gene.";
RT	J. Bacteriol. 168:1189-1196(1986).
RL	[3]
RP	SEQUENCE OF 1-14 FROM N.A.
RP	STRAIN=CPN50;
RC	MEDLINE=89039249; PubMed=2460717;
FX	Garnier T., Cole S.T.;
RA	"Studies of UV-inducible promoters from Clostridium perfringens in vivo and in vitro.";
RT	Mol. Microbiol. 2:607-614(1988).
RL	Mol. Microbiol. 2:607-614(1988).
CC	-1 FUNCTION: MAY FUNCTION AS AN IONOPHORE.
CC	-1 INDUCTION: BY UV Irradiation.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announcements).
CC	or send an email to license@ibb-sib.ch .
CC	-----
DR	EMBL; M14481; AAA98248.1; -
DR	EMBL; M32882; AAA98249.1; -
DR	PIR; A30481; A30481.
DR	InterPro; IPR003646; SH3 Bac.
DR	InterPro; IPR000834; Zn Carboxypep.
DR	Pfam; PF00246; ZnCarboxypep; 1.
DR	SMART; SM00287; SH3D; 3.
DR	Antibiotic; Bacteriocin; Plasmid.
KW	DOMAIN
FT	815 869 HYDROPHOBIC.
SO	SEQUENCE 890 AA; 96699 MW; PAE5E897IC3IC6C6 CRC64;

Query Match	63.5%	Score	33	DB	1	Length	890
Best Local	66.7%	Pred. No. 60					

Matches	6;	Conservative	1;	Mismatches	2;	Indels	0;	Gaps	0;
Qy	2	EVVVPKGMST	10						
			:						
Db	170	EVVPGGFTY	178						

Search completed: December 22, 2003, 17:42:23
Job time : 6.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:27:26 ; Search time 25.2 Seconds
(without alignments)
112.642 Million cell updates/sec

Title: US-09-909-164-5
Perfect score: 52
Sequence: 1 BEVVPXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAESP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	75.0	1044	16	Q8DIH0
2	38	73.1	387	16	Q98FX1
3	38	73.1	3472	1	Q74056
4	37	71.2	840	3	Q9URY8
5	37	71.2	1123	16	Q8BWD4
6	36	69.2	471	11	Q8RI26
7	36	69.2	484	11	Q8VD18
8	36	69.2	484	11	Q8BTX4
9	36	69.2	484	11	Q8BK35
10	35	67.3	225	10	Q40129
11	35	67.3	425	5	Q9XYK4
12	35	67.3	495	11	Q8C1D7
13	35	67.3	556	4	Q43733
14	35	67.3	583	5	Q9BH83
15	35	67.3	583	5	Q9BH45
16	35	67.3	583	5	Q815S7

17	35	67.3	670	11	Q01487	Q01487 ratfus rat
18	35	67.3	747	16	Q8PM16	Q8PM16 xanthomonas
19	35	67.3	747	16	Q8PAT2	Q8PAT2 xanthomonas
20	35	67.3	749	16	Q9PDW6	Q9PDW6 xylella fas
21	35	67.3	1902	4	Q14122	Q14122 homo sapien
22	34	65.4	156	3	Q12479	Q12479 saccharomyc
23	34	65.4	219	17	Q971S2	Q971S2 sulfolobus
24	34	65.4	252	17	Q28342	Q28342 archaeoglob
25	34	65.4	290	4	Q96MU1	Q96MU1 homo sapien
26	34	65.4	387	16	Q92MD6	Q92MD6 rhizobium m
27	34	65.4	489	4	Q81YH3	Q81YH3 homo sapien
28	34	65.4	541	16	Q98BP5	Q98BP5 rhizobium 1
29	34	65.4	544	16	Q9PD02	Q9PD02 ureaplasma
30	34	65.4	842	3	Q9URR4	Q9URR4 penicillium
31	34	65.4	899	16	Q8G415	Q8G415 bifidobacte
32	34	65.4	1049	16	Q8XT05	Q8XT05 raleconia s
33	33	63.5	143	17	Q8TX62	Q8TX62 methanopyru
34	33	63.5	166	16	Q8PP5	Q8PP5 xanthomonas
35	33	63.5	193	2	Q8VUA8	Q8VUA8 lactococcus
36	33	63.5	208	2	Q8XTQ4	Q8XTQ4 candidatur
37	33	63.5	209	16	Q8RE56	Q8RE56 fusobacteri
38	33	63.5	284	16	B74187	B74187 bynechocyst
39	33	63.5	298	10	Q9W3C0	Q9W3C0 arabidopsis
40	33	63.5	326	12	Q9QC87	Q9QC87 soil-borne
41	33	63.5	326	12	Q9Q9Q7	Q9Q9Q7 soil-borne
42	33	63.5	326	12	Q9DUJ4	Q9DUJ4 soil-borne
43	33	63.5	326	12	Q9Q9Q9	Q9Q9Q9 soil-borne
44	33	63.5	326	12	Q91DNI	Q91DNI soil-borne
45	33	63.5	326	12	Q9Q9Q5	Q9Q9Q5 soil-borne

ALIGNMENTS

Q8DIH0	Q8DIH0	PRELIMINARY;	PRT; 1044 AA.
AC	Q8DIH0		
DT	01-MAR-2003 (TREMBLrel. 23, Created)		
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)		
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)		
DE	Multidrug efflux transporter.		
GN	TLL1618.		
OS	Synechococcus elongatus (Thermosynechococcus elongatus).		
OC	Bacteria; Cyanobacteria; Chroococcales; Synechococcus.		
OX	NCBI_TaxID=32046;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BP-1;		
RX	MEDLINE=22225144; PubMed=12240834;		
RA	Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,		
RA	Watanabe A., Itiguchi M., Kawashima K., Kimura T., Kishida Y.,		
RA	Kiyokawa C., Kohara M., Matsunoto M., Matsuno A., Nakazaki N.,		
RA	Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;		
RT	*Complete genome structure of the thermophilic cyanobacterium		
RT	Thermosynechococcus elongatus BP-1.;		
RL	DNA Ref. 9123-130(2002).		
DR	EMBL; AP005374; BAC09170.1; -		
KW	Complete proteome.		
SC	SEQUENCE 1044 AA; 113205 MW; 00E9C13F0F636D2F CRC64;		
QY	Query Match	75.0%;	Score 39; DB 16; Length 1044;
DB	Beet Local Similarity	63.6%;	Pred. No. 24;
	Matches	7; Conservative	2; Mismatches
		2; Indels	0; Gaps
			0;
QY	1 BEVVPXGMSYS 11		
DB	843 BEVLPNGIGYS 853		
RESULT 2			
Q98FX1	PRELIMINARY;	PRT; 387 AA.	
ID	Q98FX1		

AC Q98FX1;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
 DE Hipurate hydrolase.
 GN MTR583.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAF03039;
 RX MEDLINE=21082330; PubMed11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Ideasa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti."
 RL DNA Res. 7:331-338 (2000).
 DR EMBL; AP003002; BAB50445.1; -
 DR InterPro: IPR002933; Peptidase_M20.
 DR Pfam; PF01546; Peptidase_M20; I.
 KW Hypochlorase; Complete proteome.
 SQ SEQUENCE 387 AA; 41180 MW; 131BF8E64306829 CRC64;

Query Match 73.1%; Score 38; DB 16; Length 387;
 Best Local Similarity 60.0%; Pred. No. 13;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 10
 DB 367 DEAIPIHGMYS 376

RESULT 3
 ID 074056 PRELIMINARY; PRT; 3472 AA.
 AC 074056;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypochlorase 367.1 kDa protein.
 OS Cenarchaeum symbiosum.
 OC Archaea; Crenarchaeota; Thermoprotei; Cenarchaeales; Cenarchaeaceae;
 OC Cenarchaeum.
 OX NCBI_TaxID=46770;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B;
 RX MEDLINE=98422450; PubMed=9748430;
 RA Schleper C., Delong E.F., Preston C.M., Feldman R.A., Wu K.Y.,
 RA Swanson R.V.;
 RT "Genomic analysis reveals chromosomal variation in natural populations
 RT of the uncultured psychrophilic archaeon Cenarchaeum symbiosum."
 RL J. Bacteriol. 180:5003-5009 (1998).
 DR EMBL; AF083072; AAC6269.1; -
 DR InterPro: IPR000515; BPD_transp.
 DR InterPro: IPR001680; WD40.
 DR Pfam; PF00400; WD40; 4.
 DR SMART; SM00320; WD40; 2.
 DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBER; 1.
 KW Hypochlorase protein; Repeat; WD repeat.
 SQ SEQUENCE 3472 AA; 367058 MW; 37F80707030F9355 CRC64;

Query Match 73.1%; Score 38; DB 1; Length 3472;
 Best Local Similarity 54.5%; Pred. No. 1; Se+02;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 11
 DB 367 DEAIPIHGMYS 376

DB 2294 EDVIPRGISFS 2304

RESULT 4
 ID Q9URY8 PRELIMINARY; PRT; 840 AA.
 AC Q9URY8;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Probable sulfatase.
 GN SPAC869.05C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetaceae; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972h-;
 RA Hunt C., Aves S., McDougall R.C., Rajandream M.A., Barrell B.G.;
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AL132779; CAB60015.1; -
 DR GeneDB; SPombe; SPAC869.05C; -
 DR InterPro: IPR002645; STAS.
 DR InterPro: IPR001902; Sulph_transp.
 DR Pfam; PF01740; STAS; 1.
 DR Pfam; PF00916; Sulfate_transp; 1.
 DR TIGRFAMs; TIGR00815; sulP; 1.
 DR PROSITE; PS00801; STAS; 1.
 SQ SEQUENCE 840 AA; 93517 MW; ED4833E162B69077 CRC64;

Query Match 71.2%; Score 37; DB 3; Length 840;
 Best Local Similarity 77.8%; Pred. No. 51;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 VVPXGMSYS 11
 DB 135 VVPXGMSYS 143

RESULT 5
 ID Q8EMD4 PRELIMINARY; PRT; 1123 AA.
 AC Q8EMD4;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE MYR 2560 paralog; 57%.
 GN MYR2710.
 OS Mycoplasma penetrans.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=28227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HF-2;
 RX MEDLINE=22354719; PubMed=12466555;
 RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
 RA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
 RT "The complete genomic sequence of Mycoplasma penetrans, an
 RT intracellular bacterial pathogen in humans."
 RL Nucleic Acids Res. 30:5283-5300 (2002).
 DR EMBL; AP004171; BAC44062.1; -
 KW Complete proteome.
 SQ SEQUENCE 1123 AA; 123636 MW; A4D707330E3DBA4C CRC64;

Query Match 71.2%; Score 37; DB 16; Length 1123;
 Best Local Similarity 70.0%; Pred. No. 70;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 EVVPXGMSYS 11
 DB 658 EVVPXGMSYS 667

RESULT 6
 08RT26 PRELIMINARY; PRT; 471 AA.
 AC 08RT26; 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DE Hypothetical 54.5 kDa protein (fragment).
 GN GLTSCR2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN NCB1_Taxid=10090;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Strausberg R.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC025810; AAH25810.1; -.
 MGD; MGI:2154441; Gltrscr2.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 471 AA; 54506 MW; E0DA685C374A9760 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 471;
 Best Local Similarity 60.0%; Pred. No. 44;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 EVVPGXMSYS 11
 DB 226 EVIPAGASYN 235

RESULT 7
 08VD18 PRELIMINARY; PRT; 484 AA.
 AC 08VD18; 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DE Similar to glioma tumor suppressor candidate region gene 2.
 GN GLTSCR2 OR AWS56441.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN NCB1_Taxid=10090;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Salivary gland;
 RA Strausberg R.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC017637; AAH17637.1; -.
 MGD; MGI:2154441; Gltrscr2.
 SQ SEQUENCE 484 AA; 55835 MW; BB845F3B4BE02A36 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
 Best Local Similarity 60.0%; Pred. No. 45;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 EVVPGXMSYS 11
 DB 239 EVIPAGASYN 248

RESULT 8
 08BTX4 PRELIMINARY; PRT; 484 AA.
 AC 08BTX4; 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DE Similar to glioma tumor suppressor candidate region gene 2.
 GN GLTSCR2 OR AWS56441.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN NCB1_Taxid=10090;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Salivary gland;
 RA Strausberg R.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC017637; AAH17637.1; -.
 MGD; MGI:2154441; Gltrscr2.
 SQ SEQUENCE 484 AA; 55835 MW; BB845F3B4BE02A36 CRC64;

DE Similar to glioma tumor suppressor CANDIDATE region gene 2
 DE protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN NCB1_Taxid=10090;
 RP SEQUENCE FROM N.A.
 RC STRAIN=NOD; TISSUE=Thymus;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 DR EMBL; AK088461; BAC40367.1; -.
 SQ SEQUENCE 484 AA; 55806 MW; B3056425B5EBCAD8 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
 Best Local Similarity 60.0%; Pred. No. 45;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 EVVPGXMSYS 11
 DB 239 EVIPAGASYN 248

RESULT 9
 08BK35 PRELIMINARY; PRT; 484 AA.
 AC 08BK35; 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DE Similar to glioma tumor suppressor CANDIDATE region gene 2
 DE protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN NCB1_Taxid=10090;
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Pituitary;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 DR EMBL; AK077341; BAC35760.1; -.
 SQ SEQUENCE 484 AA; 55792 MW; EB67949BCBE92D44 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
 Best Local Similarity 60.0%; Pred. No. 45;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 EVVPGXMSYS 11
 DB 239 EVIPAGASYN 248

RESULT 10
 040129 PRELIMINARY; PRT; 225 AA.
 AC 040129; 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DE Hypothetical 25.2 kDa protein precursor.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Solanaceae; Solanum.

OK NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF36; TISSUE=Distill;
RX MEDLINE=9537523; PubMed=7647301;
RA Milligan S.B., Gaesser C.S.;
RT "Nature and regulation of pistil-expressed genes in tomato."
RL Plant Mol. Biol. 28:691-711(1995).
DR EMBL; U20592; AAA80497.1; -.
DR InterPro; IPR002160; Kunitz_legume.
DR Pfam; PF00197; Kunitz_legume; 1.
DR PRINTS; PR00291; KUNITZINHTR.
DR ProDom; PD000891; Kunitz_legume; 1.
DR SMART; SM00452; STI; 1.
DR PROSITE; PS00283; SOYBEAN_KUNITZ; 1.
KW Hypothetical protein; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 225 UNKNOWN.
SQ SEQUENCE 225 AA; 25188 MW; 1074C261D20CFDAD CRC64;

Query Match 67.3%; Score 35; DB 10; Length 225;
Best Local Similarity 54.5%; Pred. No. 32;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
Db 32 DEVVPGKTYA 42

RESULT 11

Q9XVK4 PRELIMINARY; PRT; 425 AA.

AC Q9XVK4; 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DR 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE R10D12.10 protein.
GN R10D12.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Percy C.M.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for investigating biology."
RL Science 282:2012-2016(1998).
DR EMBL; Z81109; CAB03241.1; -.
DR WormPep; R10D12.10; C812690.
DR InterPro; IPR000719; Prot_Kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 425 AA; 49410 MW; 5D96E29B08C8E9D6 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 425;
Best Local Similarity 50.0%; Pred. No. 64;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSY 10
Db 335 EQIVPGGLQY 344

RESULT 12

Q8C1D7

ID Q8C1D7 PRELIMINARY; PRT; 495 AA.
AC Q8C1D7; 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DR 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE CDNA FLJ31891 F1S.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK028258; BAC25846.1; -.
SQ SEQUENCE 495 AA; 55358 MW; 66A22A093C114752 CRC64;

Query Match 67.3%; Score 35; DB 11; Length 495;
Best Local Similarity 63.6%; Pred. No. 75;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
Db 260 EEIVPMGISPS 270

RESULT 13

Q43733 PRELIMINARY; PRT; 556 AA.

AC Q43733; 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DR 01-JUN-1998 (TREMBlrel. 06, Last annotation update)
DE DNA binding protein (Fragment).
GN DJ451B15.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tubbey B.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z98050; CAB10847.1; -.
FT NON_TER 1 1
SQ SEQUENCE 556 AA; 59059 MW; ECB00E4033FB2528 CRC64;

Query Match 67.3%; Score 35; DB 4; Length 556;
Best Local Similarity 66.7%; Pred. No. 86;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMSYS 11
Db 244 VVPAGLTYIS 252

RESULT 14

Q9BH83 PRELIMINARY; PRT; 583 AA.

AC Q9BH83; 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DR 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Choline transporter.
GN SGT1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;

RN (1)
 RP SEQUENCE FROM N.A.
 RA Ben Mamoun C., Gluzman I.Y., Goldberg D.E.;
 RT "Plasmodium falciparum choline transporter (PfSCT1) gene."
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY007374; AAK14818.1; -
 DR EMBL; AY007373; AAK14817.1; -
 DR InterPro; IPR002123; Acyltransferase.
 DR Pfam; PF01553; Acyltransferase; 1.
 DR SMART; SM00563; Plsc; 1.
 SQ SEQUENCE 583 AA; 66918 MW; 4DF83D7530527474 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 583;
 Best Local Similarity 55.6%; Pred. No. 90;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
 DB 227 IIPVGLSYS 235

RESULT 15

Q9BHA5 PRELIMINARY; PRT; 583 AA.
 AC Q9BHA5;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Choline transporter.
 GN SCT1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NX NCBI_taxid=5833;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Ben Mamoun C., Gluzman I.Y., Goldberg D.E.;
 RT "Plasmodium falciparum choline transporter (PfSCT1) gene."
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY007372; AAK14816.1; -
 DR EMBL; AY007375; AAG17947.1; -
 DR InterPro; IPR002123; Acyltransferase.
 DR Pfam; PF01553; Acyltransferase; 1.
 DR SMART; SM00563; Plsc; 1.
 SQ SEQUENCE 583 AA; 66917 MW; 2B2BFAE3E395E049 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 583;
 Best Local Similarity 55.6%; Pred. No. 90;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
 DB 227 IIPVGLSYS 235

Search completed: December 22, 2003, 17:51:23
 Job time : 27.2667 secs

GenCore Version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: December 22, 2003, 16:41:00 ; Search time 32.4667 Seconds
(without alignments)
53.778 Million cell updates/sec

Title: US-09-909-164-6

Perfect score: 52

Sequence: 1 EEVVPKXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	96.2	11	23	ABB80521 Hepatitis C virus
2	50	96.2	11	23	ABB80522 Hepatitis C virus
3	50	96.2	11	23	ABB80525 Hepatitis C virus
4	50	96.2	11	23	ABB80526 Hepatitis C virus
5	50	96.2	11	23	ABB80559 Hepatitis C virus
6	50	96.2	11	23	ABB80563 Hepatitis C virus
7	50	96.2	11	23	ABB80564 Hepatitis C virus
8	50	96.2	11	23	ABB80565 Hepatitis C virus
9	50	96.2	11	23	ABB80566 Hepatitis C virus

1001

10	50	96.2	11	23	ABB80567 Hepatitis C virus
11	50	96.2	11	23	ABB80568 Hepatitis C virus
12	46	88.5	11	23	ABB80524 Hepatitis C virus
13	46	88.5	11	23	ABB80528 Hepatitis C virus
14	46	88.5	11	23	ABB80529 Hepatitis C virus
15	46	88.5	11	23	ABB80561 Hepatitis C virus
16	46	88.5	11	23	ABB80562 Hepatitis C virus
17	45	86.5	11	23	ABB80523 Hepatitis C virus
18	45	86.5	11	23	ABB80527 Hepatitis C virus
19	45	86.5	11	23	ABB80535 Hepatitis C virus
20	45	86.5	11	23	ABB80536 Hepatitis C virus
21	45	86.5	11	23	ABB80539 Hepatitis C virus
22	45	86.5	11	23	ABB80540 Hepatitis C virus
23	45	86.5	11	23	ABB80558 Hepatitis C virus
24	45	86.5	11	23	ABB80560 Hepatitis C virus
25	44	84.6	11	23	ABB80544 Hepatitis C virus
26	44	84.6	11	23	ABB80545 Hepatitis C virus
27	44	84.6	11	23	ABB80549 Hepatitis C virus
28	44	84.6	11	23	ABB80552 Hepatitis C virus
29	44	84.6	11	23	ABB80553 Hepatitis C virus
30	42	80.8	11	23	ABB80530 Hepatitis C virus
31	41	78.8	11	23	ABB80538 Hepatitis C virus
32	41	78.8	11	23	ABB80542 Hepatitis C virus
33	41	78.8	11	23	ABB80543 Hepatitis C virus
34	40	76.9	11	23	ABB80537 Hepatitis C virus
35	40	76.9	11	23	ABB80541 Hepatitis C virus
36	40	76.9	11	23	ABB80547 Hepatitis C virus
37	40	76.9	11	23	ABB80548 Hepatitis C virus
38	40	76.9	11	23	ABB80551 Hepatitis C virus
39	40	76.9	11	23	ABB80556 Hepatitis C virus
40	40	76.9	11	23	ABB80557 Hepatitis C virus
41	40	76.9	20	20	AAU76810 Hepatitis C virus
42	40	76.9	1022	22	ABG03621 Novel human diagno
43	40	76.9	1022	22	ABG05826 Novel human diagno
44	40	76.9	1022	22	ABG08173 Novel human diagno
45	39	75.0	11	23	ABB80546 Hepatitis C virus

ALIGNMENTS

RESULT 1					
ID	ABB80521	standard;	peptide;	11	AA.
XX	ABB80521;				
AC					
XX					
DT	08-OCT-2002	(first entry)			
XX					
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.				
XX					
KW	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketamide;				
KW	virucide.				
XX					
OS	Synthetic.				
XX					
FH	Key	Location/Qualifiers			
FT	Modified-site	1	/note= "N-terminal acetyl"		
FT	Modified-site	6	/note= "N-terminal acetyl"		
FT	Modified-site	6	/note= "N-terminal acetyl"		
FT	Modified-site	11	/note= "C-terminal amide"		
XX					
XX					
PN	WO200208251-A2.				
XX					
PD	31-JAN-2002.				
XX					
PF	19-JUL-2001; 2001WO-US23169.				
XX					
PR	21-JUL-2000; 2000US-220101P.				
XX					

CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 1 BEVVPXGMSYS 11

RESULT 4
ABB80526
ID ABB80526 standard; peptide; 11 AA.
XX
AC ABB80526;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #6.
XX
KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
PN W0200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
PS Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 1 BEVVPXGMSYS 11

RESULT 5
ABB80559
ID ABB80559 standard; peptide; 11 AA.
XX
AC ABB80559;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #39.
XX
KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 8 /note= "Oxymethionine"
FT Modified-site 11 /note= "C-terminal amide"
XX
PN W0200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
PS Claim 17; Page 65; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
|||||

[illegible]

DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #44.
XX	
XX	Hepatitis C virus; HCV, serine protease; inhibitor; alpha-ketoamide;
KM	virucide.
XX	
OS	Synthetic.
XX	
FH	Key
FT	Modified-site
FT	/note= "N-terminal acetyl]"
FT	6
FT	/notes= "Leucyl carbonyl forming keto-amide linkage with
FT	residue 7"
FT	Modified-site
FT	11
FT	/note= "C-terminal amide"
PV	WO200208251-A2.
XX	
PD	31-JAN-2002.
XX	
PF	19-JUL-2001; 2001WO-US23169.
PR	
PR	21-JUL-2000; 2000US-220101P.
XX	
PA	(CORV-) CORVAS INT INC.
PI	Llm-wilby M, Levy OE, Brunck TK,
XX	
DR	WPI, 2002-361643/39.
XX	
PT	Novel peptide compound having hepatitis C virus protease inhibitory
PT	activity useful for treating disorders associated with hepatitis C
PT	virus protease -
PS	
XX	Claim 17; Page 65; 69pp; English.
CC	
CC	The sequence represents a peptide compound of the invention having
CC	hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC	invention are alpha-ketoamide peptide analogues. The peptides have
CC	virucide activity, and are useful for treating and in the manufacture of
CC	a pharmaceutical composition comprising the peptide as an active ingredient
CC	is useful for treating disorders associated with hepatitis C virus.
XX	
SQ	Sequence 11 AA;
Query Match	96.2%; Score 50; DB 23; Length 11;
Best Local Similarity	100.0%; Pred. No. 0.0013;
Matches 11; Conservative	0; Mismatches 0; Indels 0; Gaps 0.
OY	1 EEVVPXGMSYS 11 1 EEVVPXGMSYS 11
Db	
RESULT 8	
ABB80565	
ID	ABB80565 standard; peptide; 11 AA.
XX	
AC	ABB80565;
XX	
DT	08-OCT-2002 (first entry)
XX	
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #45.
XX	
KM	Hepatitis C virus; HCV, serine protease; inhibitor; alpha-ketoamide;
XX	virucide.
XX	
OS	Synthetic.
XX	
FH	Key
FT	Location/Qualifiers
FT	Modified-site
FT	1

	FT	/note= "N-terminal acetyl"	
Modified-site	6	/note= "Norleucyl carbonyl forming keto-amide linkage with residue 7"	
FT FT			
Modified-site	11	/note= "C-terminal amide"	
PX PX			
WO200208251-A2.			
PD PD			
31-JAN-2002.			
XZ XZ			
PF PF			
19-JUL-2001; 2001WO-US23169.			
XZ XZ			
PR PR			
21-JUL-2000; 2000US-220101P.			
PA PA	(CORV-) CORVAS INT INC.		
XZ XZ			
LIM-wilby M, Levy OE, Brunck TK;			
PI PI			
WPI; 2002-361643/39.			
DZ DZ			
Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus disease -			
PS PS	Claim 17; Page 65; 69pp; English.		
The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketamide peptide analogues. The peptides have viralicidal activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus.			
SQ SQ	Sequence 11 AA:		
Query Match	96.2%; Score 50; DB 23; Length 11; Best Local Similarity 100.0%; Pred.No. 0.0013;		
Matches 11; Conservative	0; Mismatches 0; Indels 0; Gaps 0,		
QY QY	1 EEVVFXGMSYS 11 1 EEWVPFGMSYS 11		
DB DB			
RESULT 9			
ABB80566			
ID ID	ABB80566 standard, peptide; 11 AA.		
XZ XZ			
ABB80566;			
DT DT	08-OCT-2002 (first entry)		
Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #46.			
KM KM	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virocide.		
Synthetic.			
Key Location/Qualifiers			
Modified-site	1 /note= "N-terminal acetyl"		
Modified-site	6 /note= "2-aminoisobutyryl carbonyl residue forming a keto-amide linkage with residue 7"		
Modified-site	11 /note= "C-terminal amide"		
WO200208251-A2.			
31-JAN-2002.			

```

XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX L1m-w1lby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX viraluc activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus.
XX
XX Sequence 11 AA:
XX
XX Query Match 96.2%; Score 50; DB 23; Length 11;
XX Best Local Similarity 100.0%; Pred. No. 0.0013;
XX Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0
XX
XX 1 EEVVPYXGMSYS 11
XX |||||
XX 1 EEVVPYXGMSYS 11
XX
XX RESULT 10
XX ABB80567
XX ID ABB80567 standard; peptide; 11 AA.
XX
XX ABB80567;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #47.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX viralucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifier
XX Modified-site 1 /note= "N-terminal acetyl]"
XX Modified-site 6 /note= "(s,s)allochreonyl carbonyl residue forming a
XX FT FT keto-amide linkage with residue 7"
XX Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX L1m-w1lby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX

```

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX
PS Claim 17; Page 65; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
XX
Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 BEVVPXGMSYS 11
Db 1 BEVVPXGMSYS 11
XX
RESULT 11
ABB80568
ID ABB80568 standard; peptide; 11 AA.
XX
AC ABB80568;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #48.
XX
KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
OS Synthetic.
OS
FH Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Alpha-propionyl-glycyl carbonyl residue forming
FT a keto-amide linkage with residue 7"
FT Modified-site 11 /note= "C-terminal amide"
FT
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001MO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 65; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
XX
Query Match 88.5%; Score 46; DB 23; Length 11;

CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
XX
SQ Sequence 11 AA;
XX
Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 BEVVPXGMSYS 11
Db 1 BEVVPXGMSYS 11
XX
RESULT 12
ABB80524
ID ABB80524 standard; peptide; 11 AA.
XX
AC ABB80524;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.
XX
KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
OS Synthetic.
OS
FH Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001MO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
XX
Query Match 88.5%; Score 46; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.0088;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 11
| | | | |
DB 1 EEVVPXGMDYS 11

RESULT 13

ABB80528 standard; peptide; 11 AA.

AC ABB80528;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #8.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
virucide.

XX Synthetic.

FH Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C

XX PT virus protease -

XX Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 88.5%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0088;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 11
| | | | |
DB 1 EEVVPXGMDYS 11

RESULT 14
ABB80529 standard; peptide; 11 AA.

AC ABB80529;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
virucide.

XX Synthetic.

FH Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C

XX PT virus protease -

XX Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 88.5%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0088;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 11
| | | | |
DB 1 EEVVPXGMDYS 11

RESULT 15

ABB80561 standard; peptide; 11 AA.

AC ABB80561;

```

DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #41.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 8 /note= "Oxymethionine"
FT Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lin-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease -
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus.
XX
XX Sequence 11 AA;
XX
XX Query Match 88.5%; Score 46; DB 23; Length 11;
XX Best Local Similarity 90.9%; Pred. NO. 0.0088;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 EEVVPXGMSYS 11
  |||||
  |||||
Db 1 EEVVPXGMDYS 11

```

Search completed: December 22, 2003, 17:41:00
 Job time : 33.4667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:37:03 ; Search time 10.133 Seconds
(without alignment) 45.930 Million cell updates/sec

Title: US-09-909-164-6

Perfect score: 52

Sequence: 1 EEVVPXGMSYS 11

Scoring table:

BLOSUM62

Gapped 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued Patents AA:*
2: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/BACKFILES1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	65.4	947	4	US-09-228-986-73
2	34	65.4	1407	4	US-09-328-352-7885
3	33	63.5	45	2	US-08-637-759B-236
4	33	63.5	45	3	US-08-871-355A-236
5	33	63.5	45	4	US-09-201-945-236
6	33	63.5	65	6	5177197-1
7	33	63.5	410	6	5177197-1
8	33	63.5	1394	6	5177197-30
9	32	61.5	10	3	US-09-357-952-66
10	32	61.5	10	4	US-09-521-650-66
11	32	61.5	10	4	US-09-168-888-66
12	32	61.5	102	2	US-08-580-988A-23
13	32	61.5	152	2	US-08-460-694-4
14	32	61.5	152	3	US-08-460-744-4
15	32	61.5	152	3	US-07-667-711B-4
16	32	61.5	173	1	US-08-193-977-7
17	32	61.5	189	2	US-08-464-517-21
18	32	61.5	189	2	US-08-246-361A-21
19	32	61.5	189	3	US-08-463-772-21
20	32	61.5	189	5	PCT-US93-05000-21
21	32	61.5	236	2	US-08-464-517-22
22	32	61.5	236	2	US-08-246-361A-22
23	32	61.5	236	3	US-08-463-772-22
24	32	61.5	236	5	PCT-US93-05000-22
25	32	61.5	280	3	US-08-464-517-6
26	32	61.5	280	3	US-08-463-772-6
27	32	61.5	289	2	US-08-246-361A-4

28	32	61.5	289	5	PCT-US93-05000-4	Sequence 4, Appli
29	32	61.5	291	5	PCT-US93-05000-6	Sequence 6, Appli
30	32	61.5	292	2	US-08-464-517-23	Sequence 23, Appli
31	32	61.5	292	2	US-08-246-361A-6	Sequence 6, Appli
32	32	61.5	292	2	US-08-246-361A-23	Sequence 23, Appli
33	32	61.5	292	3	US-08-463-772-23	Sequence 23, Appli
34	32	61.5	292	5	PCT-US93-05000-23	Sequence 23, Appli
35	32	61.5	295	1	US-07-947-120-8	Sequence 8, Appli
36	32	61.5	295	1	US-08-472-833A-8	Sequence 8, Appli
37	32	61.5	295	2	US-08-460-694-2	Sequence 2, Appli
38	32	61.5	295	2	US-08-464-517-19	Sequence 19, Appli
39	32	61.5	295	2	US-08-464-517-20	Sequence 20, Appli
40	32	61.5	295	2	US-08-246-361A-19	Sequence 19, Appli
41	32	61.5	295	2	US-08-246-361A-20	Sequence 20, Appli
42	32	61.5	295	3	US-08-463-772-19	Sequence 19, Appli
43	32	61.5	295	3	US-08-463-772-20	Sequence 20, Appli
44	32	61.5	295	3	US-08-460-744-2	Sequence 2, Appli
45	32	61.5	295	3	US-07-667-711B-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-09-228-986-73
; Sequence 73, Application US/09228986
; Patent No. 6359198
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; FILE REFERENCE: 11000/1020
; CURRENT FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 73
; LENGTH: 947
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-228-986-73

Query Match
Best Local Similarity 65.4%; Score 34; DB 4; Length 947;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Cy 3 VVPXGMSYS 11
Db 686 VMPGISTS 694

RESULT 2
US-09-328-352-7885
; Sequence 7885, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-019A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7885
; LENGTH: 1407
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7885

Query Match
Best Local Similarity 65.4%; Score 34; DB 4; Length 1407;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

QY 2 EVVVPXGMSY 10
DB 596 EVVEGLSF 604

RESULT 3

US-08-637-759B-236
Sequence 236, Application US/08637759B
Patent No. 5876931
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/637,759B
FILING DATE: 03-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 236:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-637-759B-236

Query Match 63.5%; Score 33; DB 2; Length 45;
Best Local Similarity 60.0%; Pred. No. 6.6;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
DB 1 EEISPLGMSY 10

RESULT 4

US-08-871-355A-236
Sequence 236, Application US/08871355A
Patent No. 6015669
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street

CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/871,355A
FILING DATE: 09-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101 CON
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 236:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-871-355A-236

Query Match 63.5%; Score 33; DB 3; Length 45;
Best Local Similarity 60.0%; Pred. No. 6.6;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
DB 1 EEISPLGMSY 10

RESULT 5

US-09-201-945-236
Sequence 236, Application US/09201945
Patent No. 6342215
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/201,945
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/637,759
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:

NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPS 101
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 236:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-09-201-945-236

Query Match 63.5%; Score 33; DB 4; Length 45;
Best Local Similarity 60.0%; Pred. No. 6.6;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 EEVVPXGMSY 10
Db 1 EESISPLGMSY 10

RESULT 6
5177197-51
PATENT NO. 5177197
APPLICANT: KANZAKI, TETSUTO;OLOFSSON, ANDERS;MOREN, ANITA;
; WERNSTEDT, CHRISTER;HEILMAN, ULF;MIYAZONO, KOHEI;CLAESSON-WELSH,
; LENA;HELDIN, CARL-HENRIK
TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETA1-BINDING PROTEIN
NUMBER OF SEQUENCES: 53
CURRENT APPLICATION DATA:
; APPLICANT: Yang, Wu
; FILING DATE: 27-FEB-1990
; SEQ ID NO: 51
LENGTH: 65
5177197-51

Query Match 63.5%; Score 33; DB 6; Length 65;
Best Local Similarity 45.5%; Pred. No. 10;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 11
Db 52 KEICPGMGWYT 62

RESULT 7
5177197-1
PATENT NO. 5177197
APPLICANT: KANZAKI, TETSUTO;OLOFSSON, ANDERS;MOREN, ANITA;
; WERNSTEDT, CHRISTER;HEILMAN, ULF;MIYAZONO, KOHEI;CLAESSON-WELSH,
; LENA;HELDIN, CARL-HENRIK
TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETA1-BINDING PROTEIN
NUMBER OF SEQUENCES: 53
CURRENT APPLICATION DATA:
; APPLICANT: Yang, Wu
; FILING DATE: 27-FEB-1990
; SEQ ID NO: 1
LENGTH: 410
5177197-1

Query Match 63.5%; Score 33; DB 6; Length 410;
Best Local Similarity 45.5%; Pred. No. 81;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 11
Db 399 KEICPGMGWYT 409

RESULT 8
5177197-30
PATENT NO. 5177197
APPLICANT: KANZAKI, TETSUTO;OLOFSSON, ANDERS;MOREN, ANITA;
; WERNSTEDT, CHRISTER;HEILMAN, ULF;MIYAZONO, KOHEI;CLAESSON-WELSH,
; LENA;HELDIN, CARL-HENRIK
TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETA1-BINDING PROTEIN
NUMBER OF SEQUENCES: 53
CURRENT APPLICATION DATA:
; APPLICANT: Yang, Wu
; FILING DATE: 27-FEB-1990
; SEQ ID NO: 30
LENGTH: 1394
5177197-30

Query Match 63.5%; Score 33; DB 6; Length 1394;
Best Local Similarity 45.5%; Pred. No. 3.3e+02;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 11
Db 399 KEICPGMGWYT 409

RESULT 9
US-09-357-952-66
SEQUENCE 66, Application US/09357952
PATENT NO. 6248904
GENERAL INFORMATION:
APPLICANT: Zhang, Han-Zhong
APPLICANT: Cai, Sui Xiong
APPLICANT: Drewe, John A.
TITLE OF INVENTION: No. 6248904el Fluorescence Dyes and Their Applications for Whol
TITLE OF INVENTION: Fluorescence Screening Assays for Caspases, Peptidases, Proteas
TITLE OF INVENTION: Other Enzymes and the Use Thereof
FILE REFERENCE: 1735.0030001
CURRENT APPLICATION NUMBER: US/09/357,952
CURRENT FILING DATE: 1999-07-21
EARLIER APPLICATION NUMBER: US 60/093,642
EARLIER FILING DATE: 21-JUL-1998
NUMBER OF SEQ ID NOS: 139
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 66
LENGTH: 10
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-357-952-66

Query Match 61.5%; Score 32; DB 3; Length 10;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 EEVVPXGMSY 10
Db 1 DDIVPCMSY 10

RESULT 10
US-09-521-650-66
SEQUENCE 66, Application US/09521650
PATENT NO. 6335429
GENERAL INFORMATION:
APPLICANT: Weber, Eckard
APPLICANT: Cai, Sui Xiong
APPLICANT: Drewe, John F.W.
APPLICANT: Drewe, John A.

APPLICANT: Zhang, Han-Zhong
TITLE OF INVENTION: No. 633542961 Fluorogenic or Fluorescent Reporter Molecules and
TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence
TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the
FILE REFERENCE: 1735.0290002
CURRENT APPLICATION NUMBER: US/09/521,650
CURRENT FILING DATE: 2000-03-08
EARLIER APPLICATION NUMBER: 09/168,888
EARLIER FILING DATE: 1998-10-09
EARLIER APPLICATION NUMBER: US 60/061,582
EARLIER FILING DATE: 1997-10-10
EARLIER APPLICATION NUMBER: US 09/033,661
EARLIER FILING DATE: 1998-03-03
NUMBER OF SEQ ID NOS: 142
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 66
LENGTH: 10
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-521-650-66

Query Match 61.5%; Score 32; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 1 EEVVPXGMSY 10
Db 1 DDIVPCMSY 10

RESULT 11
US-09-168-888-66
Sequence 66, Application US/09168888
Patent No. 6342611
GENERAL INFORMATION:
APPLICANT: Weber, Eckard
APPLICANT: Cai, Sui Xiong
APPLICANT: Keana, John F.W.
APPLICANT: Drewe, John A.
APPLICANT: Zhang, Han-Zhong
TITLE OF INVENTION: No. 6342611 Fluorogenic or Fluorescent Reporter Molecules and
TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence
TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the
FILE REFERENCE: 1735.0290002
CURRENT APPLICATION NUMBER: US/09/168,888
CURRENT FILING DATE: 1998-10-09
EARLIER APPLICATION NUMBER: US 60/061,582
EARLIER FILING DATE: 1997-10-10
EARLIER APPLICATION NUMBER: US 09/033,661
EARLIER FILING DATE: 1998-03-03
NUMBER OF SEQ ID NOS: 142
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 66
LENGTH: 10
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-168-888-66

Query Match 61.5%; Score 32; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 1 EEVVPXGMSY 10
Db 1 DDIVPCMSY 10

RESULT 12
US-08-580-988A-23
Sequence 23, Application US/08580988A
Patent No. 5856161
GENERAL INFORMATION:
APPLICANT: Aggarwal et al.
TITLE OF INVENTION: Tumor Necrosis Factor
TITLE OF INVENTION: Receptor-I-Associated Protein Kinase And Methods
TITLE OF INVENTION: For Its Use
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dr. Benjamin A. Adler
STREET: 8011 Candle Lane
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77071
COMPUTER READABLE FORM:
MEDIUM TYPE: 1.44 Mo floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft word for Macintosh
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/580,988A
FILING DATE: January 3, 1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benjamin Aaron Adler, Ph.D., J.D.
REGISTRATION NUMBER: 35,423
REFERENCE/DOCKET NUMBER: D5721CIP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-777-2321
TELEFAX: 713-777-6908
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 102 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: Linear
MOLECULE TYPE: protein
DESCRIPTION: protein
HYPOTHETICAL: no
ANTI-SENSE: no
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
US-08-580-988A-23

Query Match 61.5%; Score 32; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 EEVVPXGMSY 10
Db 24 EEVFPPLMNY 33

RESULT 13
US-08-460-694-4
Sequence 4, Application US/08460694
Patent No. 5858655
GENERAL INFORMATION:
APPLICANT: Arnold, Andrew
TITLE OF INVENTION: PRAD1 Cyclin and its cDNA
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington

```

; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,694
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McConathy, Evelyn H.
; REGISTRATION NUMBER: 35,279
; REFERENCE/DOCKET NUMBER: 0609.4070002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 152 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-460-694-4

Query Match
Best Local Similarity 61.5%; Score 32; DB 2; Length 152;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
DB 20 EEVFPPLAWNY 29

RESULT 14
US-08-460-744-4
; Sequence 4, Application US/08460744
; Patent No. 6107541
; GENERAL INFORMATION:
; APPLICANT: Arnold, Andrew
; TITLE OF INVENTION: PRAD1 Cyclin and its cDNA
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,744
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McConathy, Evelyn H.
; REGISTRATION NUMBER: 35,279
; REFERENCE/DOCKET NUMBER: 0609.4070005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 152 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; MOLECULE TYPE: not relevant
; MOLECULE TYPE: not relevant
```

```

; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-460-744-4

Query Match
Best Local Similarity 61.5%; Score 32; DB 3; Length 152;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
DB 20 EEVFPPLAWNY 29

RESULT 15
US-07-667-711B-4
; Sequence 4, Application US/07667711B
; Patent No. 6110700
; GENERAL INFORMATION:
; APPLICANT: ARNOLD, ANDREW
; TITLE OF INVENTION: Prad1 Cyclin and Its cDNA
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVE., NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/667,711B
; FILING DATE: 11-MAR-1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MCPHAIL, DONALD R.
; REGISTRATION NUMBER: 35,811
; REFERENCE/DOCKET NUMBER: 0609.4070000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 152 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; US-07-667-711B-4

Query Match
Best Local Similarity 61.5%; Score 32; DB 3; Length 152;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
DB 20 EEVFPPLAWNY 29
```

Search completed: December 22, 2003, 16:43:40
Job time : 10.133 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 16:40:14 ; Search time 20.8667 Seconds
(without alignments)

98.451 Million cell updates/sec

Title: US-09-909-164-6
Perfect score: 52
Sequence: 1 EEVVPXGMSYS 11

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09C_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_NEW_PUB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10C_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	73.1	3472	12	US-10-029-120-4
2	38	73.1	3472	14	US-10-027-806-4
3	38	73.1	3472	14	US-10-034-623-4
4	38	73.1	3472	15	US-10-027-801-4
5	36	69.2	153	12	US-10-029-386-32486
6	35	67.3	236	12	US-10-029-386-32076
7	34	65.4	290	12	US-10-094-749-2565
8	34	65.4	947	15	US-10-101-464A-73
9	34	65.4	1163	12	US-10-122-067-4
10	34	65.4	1499	12	US-10-086-534-67
11	34	65.4	1499	12	US-10-122-067-2
12	33	63.5	99	15	US-10-106-698-7477
13	33	63.5	426	11	US-10-214-766-43
14	33	63.5	478	11	US-09-992-600A-108
15	33	63.5	478	11	US-09-924-340-108

16	33	63.5	478	11	US-09-746-783-184	Sequence 184, App
17	33	63.5	478	12	US-09-992-095B-108	Sequence 108, App
18	33	63.5	478	12	US-10-154-678-108	Sequence 108, App
19	33	63.5	478	12	US-09-999-570-108	Sequence 108, App
20	33	63.5	478	15	US-10-000-489-108	Sequence 108, App
21	33	63.5	478	15	US-10-000-986-108	Sequence 108, App
22	33	63.5	648	12	US-10-094-749-3010	Sequence 3010, App
23	33	63.5	653	11	US-09-820-843A-26	Sequence 26, App
24	33	63.5	1152	16	US-10-080-170-130	Sequence 130, App
25	33	63.5	1394	12	US-10-116-275-261	Sequence 261, App
26	32	61.5	10	10	US-09-947-387-66	Sequence 66, App
27	32	61.5	10	12	US-10-138-375-66	Sequence 66, App
28	32	61.5	36	12	US-10-199-820-241	Sequence 241, App
29	32	61.5	242	12	US-10-094-749-2076	Sequence 2076, App
30	32	61.5	254	9	US-09-778-927A-53	Sequence 53, App
31	32	61.5	276	12	US-10-116-275-114	Sequence 114, App
32	32	61.5	289	10	US-09-919-497-54	Sequence 54, App
33	32	61.5	289	14	US-10-024-066-2	Sequence 2, App
34	32	61.5	289	14	US-10-024-066-4	Sequence 4, App
35	32	61.5	292	12	US-10-116-275-275	Sequence 275, App
36	32	61.5	295	10	US-09-925-300-1061	Sequence 1061, App
37	32	61.5	529	9	US-09-923-304-4	Sequence 4, App
38	32	61.5	691	9	US-09-925-731-2	Sequence 2, App
39	32	61.5	691	15	US-10-101-921-4	Sequence 4, App
40	32	61.5	1377	9	US-09-815-242-10384	Sequence 10384, App
41	32	61.5	1377	12	US-10-287-274-467	Sequence 467, App
42	32	61.5	2799	14	US-10-151-735-4	Sequence 4, App
43	31	59.6	53	10	US-09-764-847-878	Sequence 878, App
44	31	59.6	53	15	US-10-092-154-878	Sequence 878, App
45	31	59.6	59	10	US-09-948-080-14	Sequence 14, App

ALIGNMENTS

RESULT 1
US-10-029-120-4
; Sequence 4, Application US/10029120
; Publication No. US20030175708A1
GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM
; FILE REFERENCE: DQOR 002A
; CURRENT APPLICATION NUMBER: US/10/029,120
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-029-120-4

Query Match 73.1%; Score 38; DB 12; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.6e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
Db 2294 EDVIPRGISFS 2304

RESULT 2
US-10-027-806-4
; Sequence 4, Application US/10027806
; Publication No. US20020160476A1
GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.

```
APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM
; FILE REFERENCE: DCCP.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-806-4

Query Match      73.1%; Score 38; DB 14; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.6e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMSYS 11
Db      2294 EDVIPRGISFS 2304

RESULT 3
US-10-034-623-4
; Sequence 4, Application US/10034623
; Publication No. US20020198365A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM
; FILE REFERENCE: DCCP.002A
; CURRENT APPLICATION NUMBER: US/10/034,623
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-034-623-4

Query Match      73.1%; Score 38; DB 14; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.6e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMSYS 11
Db      2294 EDVIPRGISFS 2304

RESULT 4
US-10-027-801-4
; Sequence 4, Application US/10027801
; Publication No. US20030054364A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM
; FILE REFERENCE: DCCP.002A
; CURRENT APPLICATION NUMBER: US/10/027,801
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-801-4

Query Match      73.1%; Score 38; DB 15; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.6e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMSYS 11
Db      2294 EDVIPRGISFS 2304

RESULT 5
US-10-029-386-32486
; Sequence 32486, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AROMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annumax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32486
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF002994.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: SWISSPROT HIT: P43363, EVALU5.006-46
US-10-029-386-32486

Query Match      69.2%; Score 36; DB 12; Length 153;
Best Local Similarity 70.0%; Pred. No. 12;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 EEVVPXGMSY 10
Db      15 KEVVPFGHSY 24

RESULT 6
US-10-029-386-32076
; Sequence 32076, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AROMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annumax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32076
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

OTHER INFORMATION: MAP TO Z98050.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.3
OTHER INFORMATION: SWISSPROT HIT: P15822, EVALUATE 1.00e-125
US-10-029-386-32076

Query Match 67.3%; Score 35; DB 12; Length 236;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 EVVFXGMSYS 11
Db 80 VVFXGLTYS 88

RESULT 7

US-10-094-749-2565
Sequence 2565, Application US/10094749
Publication No. US20030219741A1
GENERAL INFORMATION:
APPLICANT: ISOGAI, TAKAO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: OTSUKI, TETSUJI
APPLICANT: MAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHKO
APPLICANT: YOSHITAKA, TSUTOMU
APPLICANT: OTSUKA, MOTOTYUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
FILE REFERENCE: 084335/0160
CURRENT APPLICATION NUMBER: US/10/094,749
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 3381
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2565
LENGTH: 290
TYPE: PRT
ORGANISM: Homo sapiens
US-10-094-749-2565

Query Match 65.4%; Score 34; DB 12; Length 290;
Best Local Similarity 66.7%; Pred. No. 66;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMS 9
Db 35 EEIVPMGIS 43

RESULT 8
US-10-101-464A-73
Sequence 73, Application US/10101464A
Publication No. US20030046728A1
GENERAL INFORMATION:

APPLICANT: Strabala, Timothy
APPLICANT: Nieuwenhuizen, Nicolaas
APPLICANT: Higgins, Colleen M.
TITLE OF INVENTION: Compositions Isolated from Plant Cells
TITLE OF INVENTION: and their use in the Modification of Plant Cell Signaling
FILE REFERENCE: 11000.1020c2
CURRENT APPLICATION NUMBER: US/10/101,464A
CURRENT FILING DATE: 2002-03-18
PRIOR APPLICATION NUMBER: 09/704,302
PRIOR FILING DATE: 2000-11-01
PRIOR APPLICATION NUMBER: 09/228,986
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/162,866
PRIOR FILING DATE: 1999-11-01
PRIOR APPLICATION NUMBER: PCT/US00/00724
PRIOR FILING DATE: 2000-01-11
NUMBER OF SEQ ID NOS: 989
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 73
LENGTH: 947
TYPE: PRT
ORGANISM: Pinus radiata
US-10-101-464A-73

Query Match 65.4%; Score 34; DB 15; Length 947;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 EVVFXGMSYS 11
Db 686 VVPSGISYS 694

RESULT 9
US-10-122-067-4
Sequence 4, Application US/10122067
Publication No. US20030165883A1
GENERAL INFORMATION:
APPLICANT: ROY A.J. Curtis
TITLE OF INVENTION: 27091, A PHOSPHOLIPID TRANSPORTING
TITLE OF INVENTION: ATPASE MOLECULE AND USES THEREFOR
FILE REFERENCE: MP101-058P1R
CURRENT APPLICATION NUMBER: US/10/122,067
CURRENT FILING DATE: 2002-04-12
PRIOR APPLICATION NUMBER: 60/283434
PRIOR FILING DATE: 2001-04-12
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 1163
TYPE: PRT
ORGANISM: Homo sapiens
US-10-122-067-4

Query Match 65.4%; Score 34; DB 12; Length 1163;
Best Local Similarity 72.7%; Pred. No. 3.1e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
Db 133 EEVPPGSGSYS 143

RESULT 10
US-10-096-534-67
Sequence 67, Application US/10096534
Publication No. US20030166887A1
GENERAL INFORMATION:
APPLICANT: The Brigham and Women's Hospital, Inc.
APPLICANT: Yates, Karen
APPLICANT: Mizuno, Shuichi
APPLICANT: Glowacki, Julie
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF SKELETAL DEGENERATION CONDITIONS

```
FILE REFERENCE: B0801/7244/KA/ERP
CURRENT APPLICATION NUMBER: US/10/096,534
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: US 60/274,980
PRIOR FILING DATE: 2001-03-12
NUMBER OF SEQ ID NOS: 79
SOFTWARE: PatentIn version 3.0
SEQ ID NO 67
LENGTH: 1499
TYPE: PRT
ORGANISM: Homo sapiens
US-10-096-534-67

Query Match
Best Local Similarity 65.4%; Score 34; DB 12; Length 1499;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
Db 469 EEVPRGGSVS 479

RESULT 11
US-10-122-067-2
Sequence 2, Application US/10122067
Publication No. US20030165883A1
GENERAL INFORMATION:
APPLICANT: ROY A.J. CURTIS
TITLE OF INVENTION: A PHOSPHOLIPID TRANSPORTING
FILE REFERENCE: MPI01-058P1R
CURRENT APPLICATION NUMBER: US/10/122,067
CURRENT FILING DATE: 2002-04-12
PRIOR APPLICATION NUMBER: 60/283434
PRIOR FILING DATE: 2001-04-12
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 1499
TYPE: PRT
ORGANISM: Homo sapiens
US-10-122-067-2

Query Match
Best Local Similarity 65.4%; Score 34; DB 12; Length 1499;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
Db 469 EEVPRGGSVS 479

RESULT 12
US-10-106-698-7477
Sequence 7477, Application US/10106698
Publication No. US20030109690A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
FILE REFERENCE: PA005P1
CURRENT APPLICATION NUMBER: US/10/106,698
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 7477
LENGTH: 99
TYPE: PRT
```

```
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (3)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (13)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (14)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (39)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (40)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (57)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (58)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (63)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (65)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (82)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (93)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-7477

Query Match
Best Local Similarity 63.5%; Score 33; DB 15; Length 99;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
Db 73 LVPGIGISYS 81

RESULT 13
US-10-214-766-43
Sequence 43, Application US/10214766
Publication No. US20030084473A1
GENERAL INFORMATION:
APPLICANT: Gocal, Greg
TITLE OF INVENTION: NON-TRANSGENIC HERBICIDE RESISTANT PLANTS
FILE REFERENCE: CA1138
CURRENT APPLICATION NUMBER: US/10/214,766
CURRENT FILING DATE: 2002-08-09
PRIOR APPLICATION NUMBER: US 60/311,734
PRIOR FILING DATE: 2001-08-09
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.2
SEQ ID NO 43
LENGTH: 426
TYPE: PRT
ORGANISM: Vibrio cholerae
US-10-214-766-43

Query Match
Best Local Similarity 63.5%; Score 33; DB 15; Length 426;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 10
Db 223 EFVTPAGOSY 232
```

RESULT 14

US-09-992-600A-108
; Sequence 108, Application US/09992600A
; Publication No. US20030027161A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US4.DIV
; CURRENT APPLICATION NUMBER: US/09/992,600A
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: JPatent
; SEQ ID NO: 108
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-992-600A-108

Query Match

Best Local Similarity 63.5%; Score 33; DB 11; Length 478;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 11

Db 239 EVAPAGASYN 248

RESULT 15

US-09-924-340-108
; Sequence 108, Application US/09924340
; Publication No. US20030027248A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US2.REG
; CURRENT APPLICATION NUMBER: US/09/924,340
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO: 108
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-924-340-108

Query Match

Best Local Similarity 63.5%; Score 33; DB 11; Length 478;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 11
Db 239 EVAPAGASYN 248

Search completed: December 22, 2003, 17:32:40
Job time : 20.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 17:24:36 ; Search time 9.06667 Seconds
(without alignments)

116.675 Million cell updates/sec

Title: US-09-909-164-6
Perfect score: 52
Sequence: 1 EEVVPXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	73.1	3472	2	T31308 hypothetical 367K
2	37	71.2	840	2	T39116 probable sulfate p
3	37	71.2	877	2	T40413 sulfate permease -
4	36	69.2	102	2	A42452 VI protein - tobac
5	36	69.2	1498	2	B97355 DNA segregation AT
6	35	67.3	225	2	S57810 hypothetical prote
7	35	67.3	425	2	T24111 hypothetical prote
8	35	67.3	670	2	S22293 zinc finger protei
9	35	67.3	749	2	H82691 DNA-binding protei
10	35	67.3	2717	2	A34203 topoisomerase IV a
11	34	65.4	156	2	S54619 hypothetical prote
12	34	65.4	252	2	H69491 cell division inhi
13	34	65.4	544	2	C82900 probable ABC sub
14	34	65.5	94	2	I40758 50S ribosomal prot
15	33	63.5	116	2	E90544 hypothetical prote
16	33	63.5	165	2	D81374 hypothetical prote
17	33	63.5	253	2	C81374 hypothetical prote
18	33	63.5	259	2	T34536 hypothetical prote
19	33	63.5	284	2	S75817 beta-ketocacyl-ACP
20	33	63.5	298	2	T47670 3-phosphoshikimate
21	33	63.5	368	2	F72281 cdc37 protein - fi
22	33	63.5	426	2	D82163 iron(III) ABC tran
23	33	63.5	466	2	T43653 bacteriocin BCN5 -
24	33	63.5	653	2	D83352 ATP-dependent DNA
25	33	63.5	890	2	A30481 conserved hypotet
26	33	63.5	1028	2	AE3286 transforming growt
27	33	63.5	1152	2	D87046 DNA-directed RNA p
28	33	63.5	1394	2	A35626
29	33	63.5	1401	2	G82336

30	33	63.5	1548	2	T04456 hypothetical prote
31	33	63.5	1712	2	A38261 making protein pr
32	32	61.5	84	2	E97333 hypothetical prote
33	32	61.5	175	2	P00616 transport protein
34	32	61.5	223	2	T01457 rho protein GDP-di
35	32	61.5	279	2	B72481 hypothetical prote
36	32	61.5	288	2	JC4011 cyclin D2 - rat
37	32	61.5	288	2	I58372 cyclin D2 - mouse
38	32	61.5	289	2	A41964 cyclin D2 - rat
39	32	61.5	289	2	A42822 cyclin D2 - human
40	32	61.5	291	2	S57922 cyclin D1 - Africa
41	32	61.5	291	2	S57925 cyclin D2 - Africa
42	32	61.5	291	2	JC4579 cyclin D2 - chicken
43	32	61.5	291	2	S62730 cyclin D1 - zebra
44	32	61.5	292	2	B42832 cyclin D3 - human
45	32	61.5	295	2	A38977 cyclin D1 - human

ALIGNMENTS

RESULT 1

hypothetical 367K protein - Cenarchaeum symbiosum
C.Species: Cenarchaeum symbiosum
C.Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
R.Schleper, C. Delong, E.F.; Preston, C.M.; Feldman, R.A.; Wu, K.Y.; Swanson, R.V.
J. Bacteriol. 180, 5003-5009, 1998
A.Title: Genomic analysis reveals chromosomal variation in natural populations of the
A.Reference number: Z20994; MUID:98422450; PMID:9748430
A.Accession: T31308
A.Status: preliminary; translated from GB/EMBL/DBD
A.Molecule type: DNA
A.Residues: 1-3472 <SCH>
A.Cross-references: EMBL:AF083072; NID:g3599393; PID:g3599394; PIDN:AAC62699.1
C.Superfamily: Cenarchaeum symbiosum hypothetical 367K protein

Query Match 73.1%; Score 38; DB 2; Length 3472;
Best Local Similarity 54.5%; Pred. NO. 59;

Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
Db 2294 EDVIRPGISFS 2304

RESULT 2

T39116 probable sulfate permease - fission yeast (Schizosaccharomyces pombe)

C.Species: Schizosaccharomyces pombe
C.Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C.Accession: T39116

R.Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL data library, November 1999

A.Reference number: Z21829

A.Accession: T39116

A.Status: preliminary; translated from GB/EMBL/DBD

A.Molecule type: DNA

A.Residues: 1-840 <HUN>

A.Cross-references: EMBL:AL132779; PIDN:CA860015.1; GSPDB:GN00066; SPDB:SPAC869.05c

C.Genetics:

A.Gene: SPDB:SPAC869.05c

A.Map position: 1

Query Match 71.2%; Score 37; DB 2; Length 840;
Best Local Similarity 77.8%; Pred. NO. 21;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVVPXGMSYS 11
Db 135 VVPGQMSYA 143

RESULT 3

T40413 sulfate permease - fission yeast (*Schizosaccharomyces pombe*)

C:Species: *Schizosaccharomyces pombe*

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T40413

R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Jimenez Martinez, J.

A:Reference number: 221926

A:Accession: T40413

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-877 <LYN>

A:Cross-references: EMBL:AL031261; PIDN:CAA20298.1; GSPDB:GN00067; SPDB:SPBC3H7.02

A:Experimental source: strain 972h-; cosmid c3H7

A:Genetics:

A:Gene: SPDB:SPBC3H7.02

A:Map position: 2

A:Query Match

Best Local Similarity 77.8%; Pred. No. 22;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 EVVPXGMSYS 11

DB 148 VVPGGMSYA 156

RESULT 4

A42452

VI protease - tobacco yellow dwarf virus (strain Australia)

C:Species: tobacco yellow dwarf virus

C:Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999

C:Accession: A42452

R:Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.

A:Reference number: A42452; PMID:1546458

A:Accession: A42452

A:Molecule type: DNA

A:Residues: 1-102 <MOR>

A:Cross-references: GB:M81103; NID:935283; PIDN:AAA47947.1; PID:935284

A:Query Match

Best Local Similarity 69.2%; Score 36; DB 2; Length 102;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 11

DB 7 QVVPXGMSYS 16

RESULT 5

B97355

DNA segregation ATPase, FleK/SpoIIIE family, YUKA B. subtilis ortholog [imported] - Clo

C:Species: *Clostridium acetobutylicum*

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: B97355

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

A:Reference number: A96900; PMID:21359325; PMID:21359325

A:Accession: B97355

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1498 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK81629.1; PID:G15026014; GSPDB:GN00168

A:Experimental source: *Clostridium acetobutylicum* ATCC824

A:Genetics:

A:Gene: CAC3709

Query Match 69.2%; Score 36; DB 2; Length 1498;

Best Local Similarity 60.0%; Pred. No. 63;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVPXGMSY 10

DB 1276 EQKIPGMSY 1285

RESULT 6

S57810

hypothetical protein precursor (clone TPP11) - tomato

C:Species: *Lycopersicon esculentum* (tomato)

C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000

C:Accession: S57810

R:Milligan, S.B.; Gasser, C.S.

A:Reference number: S57808; PMID:95375233; PMID:7647301

A:Accession: S57810

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-225 <MTL>

A:Cross-references: EMBL:U20592; NID:9924625; PIDN:AAA0497.1; PID:9924626

A:Experimental source: plant Kunlun-type proteinase inhibitor

A:Query Match

Best Local Similarity 67.3%; Score 35; DB 2; Length 225;

Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVPXGMSYS 11

DB 32 DEVPNGKTYA 42

RESULT 7

T24111

hypothetical protein R10D12.10 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T24111

R:Percy, C.

A:Reference number: T24111

A:Accession: T24111

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-425 <MTL>

A:Cross-references: EMBL:Z81109; PIDN:CA03241.1; GSPDB:GN00023; CESP:R10D12.10

A:Experimental source: clone R10D12

A:Genetics:

A:Gene: CESP:R10D12.10

A:Map position: 5

A:Introns: 23/3; 56/3; 113/3; 257/2

Query Match 67.3%; Score 35; DB 2; Length 425;

Best Local Similarity 50.0%; Pred. No. 26;

Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVPXGMSY 10

DB 335 EQIVPGLOY 344

RESULT 8

S22293

zinc finger protein AT-BP2 - rat (fragment)

C:Species: *Rattus norvegicus* (Norway rat)

C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 20-Sep-1999

C:Accession: S22293; 178656

R:Michelson, C.; Traboni, C.; Cortese, R.

A:Reference number: S22293; PMID:95375233; PMID:7647301

A:Accession: S22293

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1498 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK81629.1; PID:G15026014; GSPDB:GN00168

A:Experimental source: *Clostridium acetobutylicum* ATCC824

A:Genetics:

A:Gene: CAC3709

A>Title: Isolation of two cDNAs encoding zinc finger proteins which bind to the alpha 1-A;Reference number: 158280; MUID:91187610; PMID:1901405
 A;Accession: S22293
 A;Status: nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 1-670 <MIT>
 A;Cross-references: EMBL:X54250; NID:957519; PIDN:CAA38151.1; PID:957520
 A;Note: the authors did not translate the codon for residue 1
 C;Superfamily: HIV-SP2 enhancer-binding protein
 C;Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 670;
 Best Local Similarity 66.7%; Pred. No. 43;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 EVVFXGMSYS 11
 ||| |||
 Db 376 VVPAGLTYS 384

RESULT 9
 H82691
 topoisomerase IV subunit XPI353 [imported] - Xylella fastidiosa (strain 9a5c)
 C;Species: Xylella fastidiosa
 C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C;Accession: H82691
 R;Anonymous: The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences
 A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A;Reference number: A82515; MUID:20365717; PMID:10910347
 A;Note: for a complete list of authors see reference number A59328 below
 A;Accession: H82691
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-749 <STM>
 A;Cross-references: GB:AE003967; GB:AE003849; NID:9106347; PIDN:AAF84162.1; GSPDB:GN001
 A;Experimental source: strain 9a5c
 R;Stimpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
 as-Neto, E.; Docena, C.; El-Dorty, H.; Facincani, A.F.; Ferreira, A.J.S.
 submitted to Genbank, June 2000

Query Match 67.3%; Score 35; DB 2; Length 749;
 Best Local Similarity 77.8%; Pred. No. 48;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVFXGMSY 10
 ||| |||
 Db 526 EVDPGMSY 534

RESULT 10
 A34203
 DNA-binding protein PRDII-BP1 - human
 N;Alternate names: major histocompatibility complex enhancer-binding protein 1
 C;Species: Homo sapiens (man)
 C;Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 20-Sep-1999
 C;Accession: A34203; A34779
 R;Fan, C.M.; Maniatis, T.
 Genes Dev. 4, 29-42, 1990

A>Title: A DNA-binding protein containing two widely separated zinc finger motifs that
 A;Reference number: A34203; MUID:90169514; PMID:2106471
 A;Accession: A34203
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-2717 <FAN>
 A;Cross-references: EMBL:X51435; NID:938017; PIDN:CAA35798.1; PID:938018
 R;Baldwin Jr., A.S.; LeClair, K.P.; Singh, H.; Sharp, P.A.
 Mol. Cell. Biol. 10, 1406-1414, 1990
 A>Title: A large protein containing zinc finger domains binds to related sequence elem
 A;Reference number: A34779; MUID:90205817; PMID:2108316
 A;Accession: A34779
 A;Status: preliminary; nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 801-1072; N:1074-1168; K:1170-1225; V:1227-1434; N:1436-1607; I:1609-
 A;Cross-references: GB:M32019
 C;Superfamily: HIV-SP2 enhancer-binding protein
 C;Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 2717;
 Best Local Similarity 66.7%; Pred. No. 1,9e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
 ||| |||
 Db 2405 VVPAGLTYS 2413

RESULT 11
 S54619
 hypothetical protein YOR013w - yeast (Saccharomyces cerevisiae)
 N;Alternate names: hypothetical protein O2612; hypothetical protein YOL303.3
 C;Species: Saccharomyces cerevisiae
 C;Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
 C;Accession: S54619; S66879
 R;de Haan, M.; Maarse, A.C.; Grievell, L.A.
 submitted to the EMBL Data Library, May 1995
 A;Reference number: S54617
 A;Accession: S54619
 A;Molecule type: DNA
 A;Residues: 1-156 <DEH>
 A;Cross-references: EMBL:X87331; NID:91041652; PIDN:CAA60762.1; PID:9829123
 R;de Haan, M.; Grievell, L.A.; Maarse, A.C.
 submitted to the Protein Sequence Database, July 1996

Query Match 65.4%; Score 34; DB 2; Length 156;
 Best Local Similarity 66.7%; Pred. No. 14;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVFXGMSY 10
 ||| |||
 Db 50 EVVPLGMDY 58

RESULT 12
 H69491
 cell division inhibitor (mind-2) homolog - Archaeoglobus fulgidus
 C;Species: Archaeoglobus fulgidus
 C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 26-Aug-1999
 C;Accession: H69491
 R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Doder
 ; Pletschmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.I
 Glodok, A.; Zhou, L.; Overbeek, R.; Cocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997
A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Mosee, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: H69491
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-252 <RES>
A:Cross-references: GB:AE000970; GB:AE000782; NID:G2689293; PIDN:AA89318.1; PID:G264860
C:Superfamily: cell division inhibitor mind

Query Match 65.4%; Score 34; DB 2; Length 252;
Best Local Similarity 75.0%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 EVVFXGMS 9
|||
Db 81 EVVFXGMS 86

RESULT 13

C82900
Probable ABC substrate-binding protein, iron UJ359 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: C82900
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Caswell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mif
A:Reference number: A82870
A:Accession: C82900
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-544 <GLA>
A:Cross-references: GB:AE002133; GB:AF222894; NID:G6899339; PIDN:AAF30768.1; GSPDB:GN001
C:Experimental source: serovar 3; biovar 1
C:Genetic: A:Gene: ABCSdp-5; UJ359
A:Genetic code: SGC3

Query Match 65.4%; Score 34; DB 2; Length 544;
Best Local Similarity 70.0%; Pred. No. 55;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EEVVPXGMSY 10
|||
Db 135 EEVVPXGMSY 144

RESULT 14

140758
hypothetical protein 1 - Campylobacter jejuni (fragment)
C:Species: Campylobacter jejuni
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
C:Accession: 140758; S47317
R:Hani, E.K.; Chan, V.L.
J. Bacteriol. 177, 2396-2402, 1995
A:Title: Expression and characterization of Campylobacter jejuni benzoylglycine amidohyd
A:Reference number: 140758; MUID:95247673; PMID:7730270
A:Accession: 140758
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-94 <RES>
A:Cross-references: EMBL:Z36940; NID:G535805; PIDN:CAA85392.1; PID:G535806

Query Match 63.5%; Score 33; DB 2; Length 94;
Best Local Similarity 55.6%; Pred. No. 14;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 EVVFXGMSY 10
|||
Db 26 DIFPFGMSY 34

RESULT 15

E90544
50S ribosomal protein L20 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: E90544
R:Chandaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moser, I.
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pul
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: E90544
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-116 <KUR>
A:Cross-references: GB:AL445566; PID:G14089674; PIDN:CAC13434.1; GSPDB:GN00153
A:Experimental source: strain UAB CTIP
C:Genetic: A:Gene: MYPV 2610
A:Genetic code: SGC3
C:Superfamily: Escherichia coli ribosomal protein L20

Query Match 63.5%; Score 33; DB 2; Length 116;
Best Local Similarity 77.8%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 VVPXGMSYS 11
|||
Db 68 VVPXGMSYS 76

Search completed: December 22, 2003, 17:44:52
Job time : 9.0667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:43:51 ; Search time 4.6 Seconds

(without alignments)
112.455 Million cell updates/sec

Title: US-09-909-164-6

Percent score: 52

Sequence: 1 EEVVPKXGMSYS 11

Scoring table: BLOSUM62

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	73.1	1058	1	CARB_FUSNN
2	37	71.2	877	1	SULH_SCHPO
3	36	69.2	102	1	YILK_TYDPA
4	36	69.2	1498	1	YIA9_CLOAB
5	35	67.3	2717	1	ZBP1_HUMAN
6	34	65.4	788	1	CT14_NEUR
7	34	65.4	1499	1	A10C_HUMAN
8	33	63.5	116	1	RL20_MYCP
9	33	63.5	165	1	YJ49_ARCFU
10	33	63.5	253	1	Y990_CAME
11	33	63.5	280	1	CTK3_MOUSE
12	33	63.5	426	1	AKO4_VIBCH
13	33	63.5	466	1	CC37_SCHPO
14	33	63.5	478	1	GSR2_HUMAN
15	33	63.5	890	1	BCN5_CLOPE
16	33	63.5	1389	1	LTBS_MOUSE
17	33	63.5	1394	1	LTBS_HUMAN
18	33	63.5	1401	1	RPOC_VIBCH
19	33	63.5	1595	1	LTBL_HUMAN
20	33	63.5	1712	1	LTBL_RAT
21	33	63.5	1713	1	LTBL_MOUSE
22	32	61.5	288	1	CGD2_RAT
23	32	61.5	289	1	CGD2_HUMAN
24	32	61.5	289	1	CGD2_MOUSE
25	32	61.5	291	1	CGD1_BRASE
26	32	61.5	291	1	CGD1_XENLA
27	32	61.5	291	1	CGD2_CHICK
28	32	61.5	291	1	CGD2_XENLA
29	32	61.5	292	1	CGD3_CHICK
30	32	61.5	292	1	CGD3_HUMAN
31	32	61.5	295	1	CGD1_HUMAN
32	32	61.5	295	1	CGD1_MOUSE
33	32	61.5	295	1	CGD1_RAT

34	32	61.5	427	1	TOLB_HAEIN	P44677 haemophilus
35	32	61.5	529	1	ENP3_HUMAN	O75355 homo sapien
36	32	61.5	691	1	S216_HUMAN	O95616 homo sapien
37	32	61.5	719	1	GSP_GRIPI	P90518 cichthidia f
38	32	61.5	726	1	PRTF_HSV6U	P52384 human herpe
39	32	61.5	759	1	SCCT1_YEAST	P32784 saccharomyc
40	32	61.5	920	1	EDD_RAT	O62671 ratius norv
41	32	61.5	993	1	VIA_TAV	P28931 tomato aape
42	32	61.5	1377	1	RHSA_ECOLI	P16916 escherichia
43	32	61.5	1397	1	RHSC_ECOLI	P16917 escherichia
44	32	61.5	1411	1	RHSE_ECOLI	O95071 homo sapien
45	32	61.5	2799	1	EDD_HUMAN	

ALIGNMENTS

RESULT 1
CARB_FUSNN STANDARD; PRT; 1058 AA.
ID CARB_FUSNN
AC O8RG86;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) [Carbamoyl]-
DE phosphate synthetase ammonia chain).
GN CARB OR F00422
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_Taxid=76856;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldstein E., Bernal A.,
RA Larsen N., D'Souza M., Malinas T., Pusch G., Haasekorn R.,
RA Forstein M., Kyrides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT J. Bacteriol. 184:2005-2018(2002).
CC -1- CAPTALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -1- COPACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -1- PATHWAY: Arginine biosynthesis.
CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL, A010554; AA094625.1; ALT_INIT.
CC HANAP; MF_01210; -; 1.
DR InterPro: IPR006275; CARB_L_glu.
DR InterPro: IPR005483; CPase_L.
DR InterPro: IPR005479; CPase_L_D2.
DR InterPro: IPR005480; CPase_L_D3.
DR InterPro: IPR005481; CPase_L_N.
DR InterPro: IPR004362; MGS_Like.
DR Pfam: PF00289; CPase_L_Chain; 2.
DR Pfam: PF02786; CPase_L_D2; 2.
DR Pfam: PF02787; CPase_L_D3; 1.

DR Pfam; PF00142; MGS; 1.
 DR PRINTS; PRO00098; CPASASE.
 DR TIGRFAMs; TIGR01369; CPASase1_lrg. 1.
 DR PROSITE; PS00866; CPASASE_1; 2.
 DR PROSITE; PS00867; CPASASE_2; 2.
 KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 ATP-binding; Manganese; Complete proteome.
 FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 546 546 OLIGOMERIZATION DOMAIN.
 FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
 FT REPEAT 1 546
 FT REPEAT 547 1058
 FT NP_BIND 153 210 ATP (POTENTIAL).
 FT NP_BIND 302 352 ATP (POTENTIAL).
 FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
 FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
 FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
 FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
 FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
 SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1E39F CRC64;

Query Match 73.1%; Score 38; DB 1; Length 1058;
 Best Local Similarity 60.0%; Pred. No. 6.3;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPGMSYS 11
 DB 190 EIVNGGLNYS 199

RESULT 2

ID SUH_SCHPO STANDARD; PRT; 877 AA.
 AC 074377;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable sulfate permease C3H7.02.
 GN SPBC3H7.02.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; Pubmed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Segures J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holtroyd S., Hornby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Mounle S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Voiclaert G., Aert R., Robben J., Grymporter B.,
 RA Weltjens I., Vanstrele E., Rieger M., Schaefer M., Wheller-Auer S.,
 RA Gabel K., Fuchs M., Fritze C., Holzer E., Moesli D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leinrich H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleau V., Motier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,
 RA Dada R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;

RT "The genome sequence of Schizosaccharomyces pombe";

RL Nature 415:871-880(2002).

CC -1- FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE INTO THE CELL (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE SLC26A FAMILY OF TRANSPORTERS.
 CC -1- SIMILARITY: Contains 1 STAS domain.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AL031261; CAA20298.1; -.

DR PIR; T40413; T40413.

DR GeneDB Spombe; SPBC3H7.02; -.

DR InterPro; IPR002645; STAS.

DR InterPro; IPR001902; Sulph_transpt.

DR Pfam; PF01740; STAS; 1. Sulfate_transp; 1.

DR TIGRFAMs; TIGR00815; sulp; 1.

DR PROSITE; PS01130; SLC26A; 1.

DR PROSITE; PS50801; STAS; 1.

KW Transport; Transmembrane.

FT TRANSMEM 133 153 POTENTIAL.

FT TRANSMEM 161 181 POTENTIAL.

FT TRANSMEM 186 206 POTENTIAL.

FT TRANSMEM 221 241 POTENTIAL.

FT TRANSMEM 243 263 POTENTIAL.

FT TRANSMEM 292 312 POTENTIAL.

FT TRANSMEM 329 349 POTENTIAL.

FT TRANSMEM 384 404 POTENTIAL.

FT TRANSMEM 424 444 POTENTIAL.

FT TRANSMEM 461 481 POTENTIAL.

FT TRANSMEM 484 504 POTENTIAL.

FT TRANSMEM 518 538 POTENTIAL.

FT TRANSMEM 543 563 POTENTIAL.

FT DOMAIN 594 747 STAS.

SQ SEQUENCE 877 AA; 96373 MW; 56995A8493371E43 CRC64;

Query Match 71.2%; Score 37; DB 1; Length 877;
 Best Local Similarity 77.8%; Pred. No. 8.4;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPGMSYS 11

DB 148 VVPGMSYA 156

RESULT 3

ID Y1LK TYDVA STANDARD; PRT; 102 AA.
 AC P31619;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-1993 (Rel. 27, Last annotation update)
 DE Hypothetical 11.2 kDa protein.
 GN Y1.
 OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
 OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
 OX NCBI_TaxID=31599;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2188538; Pubmed=1546458;
 RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;
 RT "The nucleotide sequence of the infectious cloned DNA component of
 RT tobacco yellow dwarf virus reveals features of geminiviruses
 RT infecting monocotyledonous plants.";
 RL Virology 187:633-642(1992).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M81103; AAA47947.1; -
 DR PIR; A42452; A42452.
 DR InterPro: IPR002621; Gemini_mov.
 DR Pfam; PF01708; Gemini_mov; 1.
 DR Hypothetical protein.
 SQ SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF55B67 CRC64;

Query Match 69.2%; Score 36; DB 1; Length 102;
 Best Local Similarity 60.0%; Pred. No. 1.4;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPGXMSYS 11
 DB 7 QVVPSCINYS 16

RESULT 4
 Y1A9_CLOAB STANDARD; PRT; 1498 AA.
 ID 004351;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hypothetical protein CAC3709.
 GN CAC3709.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 NCBI_TaxID=1488;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=2159325; PubMed=1466286;
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum";
 RL J. Bacteriol. 183:4823-4838(2001).
 [2]
 RP SEQUENCE OF 1-108 FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=93273706; PubMed=8501044;
 RA Sauer U., Duerre P.;
 RT "Sequence and molecular characterization of a DNA region encoding a
 RT small heat shock protein of Clostridium acetobutylicum";
 RL J. Bacteriol. 175:3394-3400(1993).
 [3]
 CC -1- SIMILARITY: Contains 2 Pfsk/SpoIIIE domains.
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
 CC in positions 76 and 106.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AE007866; AAK81629.1; -
 DR EMBL; X65276; GAA46379.1; ALT_FRAME.
 DR PIR; B97355; B97355.
 DR InterPro: IPR002543; Pfsk_SpoIIIE.
 DR Pfam; PF01580; Pfsk_SpoIIIE; 2.

DR PROSITE; P550901; FTSK_SPOIIIE; 2.
 KW Hypothetical protein; ATP-binding; Complete proteome; Repeat.
 FT DOMAIN 655 857 FTSK/SPOIIIE 1.
 FT NP BIND 675 682 ATP (POTENTIAL).
 FT DOMAIN 1001 1188 FTSK/SPOIIIE 2.
 SQ SEQUENCE 1498 AA; 168968 MW; FF42037A335A9649 CRC64;

Query Match 69.2%; Score 36; DB 1; Length 1498;
 Best Local Similarity 60.0%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
 DB 1276 EOKIPMGMSY 1285

RESULT 5
 ZEP1_HUMAN STANDARD; PRT; 2717 AA.
 ID P15822;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein 40 (Human immunodeficiency virus type I enhancer-
 DE binding protein 1) (HIV-Ep1) (Major histocompatibility complex binding
 DE protein 1) (MBP-1) (Positive regulatory domain II binding factor 1)
 DE (PRO1-BF1).
 GN HIVP1 OR ZNF40.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90169514; PubMed=2106471;
 RA Fan C.M., Maniatis T.;
 RT "A DNA-binding protein containing two widely separated zinc finger
 RT motifs that recognize the same DNA sequence";
 RL Genes Dev. 4:29-42(1990).
 [2]
 RP STRUCTURE BY NMR OF 2113-2142.
 RX MEDLINE=9106433; PubMed=2248949;
 RA Omichinski J.G., Clore G.M., Appella E., Sakaguchi K.,
 RA Gronenborn A.M.;
 RT "High-resolution three-dimensional structure of a single zinc finger
 RT from a human enhancer binding protein in solution";
 RL Biochemistry 29:9324-9334(1990).
 [3]
 RP STRUCTURE BY NMR OF 2087-2142.
 RX MEDLINE=92232684; PubMed=1567844;
 RA Omichinski J.G., Clore G.M., Roblen M., Sakaguchi K., Appella E.,
 RA Gronenborn A.M.;
 RT "High-resolution solution structure of the double Cys2His2 zinc
 RT finger from the human enhancer binding protein MBP-1";
 RL Biochemistry 31:3907-3917(1992).
 CC -1- FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE
 CC 5'-GGGACTTC-3' WHICH IS FOUND IN THE ENHANCER ELEMENTS OF
 CC NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV.
 CC IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS
 CC OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I
 CC MHC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT
 CC IN T-CELL ACTIVATION.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- INDUCTION: BY MITOGEN AND PHORBOL ESTER.
 CC -1- DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY
 CC SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH
 CC ZINC-FINGER IN-BETWEEN.
 CC -1- SIMILARITY: STRONG, TO HIVP2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL, X51435, CA935798.1; -

DR PIR, A34203; A34203.

DR PDB, 3ZNF; 15-JAN-92.

DR PDB, 4ZNF; 15-JAN-92.

DR PDB, 1BBO; 31-OCT-93.

DR TRANSFAC; T00497; -

DR Genew; HGNC:4920; HIVEP1.

DR MIM; 194540; -

DR GO; GO:0005634; C:nucleus; TAS.

DR GO; GO:0003677; F:DNA binding activity; TAS.

DR InterPro; IPR007087; Znf C2H2.

DR Pfam; PF00096; ZF-C2H2; 5.

DR SMART; SM00355; ZNF_C2H2; 4.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.

DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.

KM Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;

KW Nuclear protein; Repeat; 3d-structure.

FT ZN_FING 406 428 C2H2-TYPE.

FT ZN_FING 434 456 C2H2-TYPE.

FT ZN_FING 958 981 C2HC-TYPE (POTENTIAL).

FT ZN_FING 2087 2109 C2H2-TYPE.

FT ZN_FING 2115 2139 C2H2-TYPE.

FT DOMAIN 803 806 POLY-SER.

FT STRAND 2088 2088

FT TURN 2090 2092

FT STRAND 2095 2095

FT HELIX 2099 2108

FT TURN 2109 2109

FT STRAND 2115 2116

FT STRAND 2123 2124

FT HELIX 2127 2135

SO SEQUENCE 2177 AA; 297217 MW; D4SD3CA951FEA561 CRC64;

Query Match 67.3%; Score 35; DB 1; Length 2717;

Best Local Similarity 66.7%; Pred. No. 73;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11

DB 2405 VVPAGLTVS 2413

RESULT 6

CY14_NEUCR STANDARD; PRT; 788 AA.

AC P23622;

DT 01-NOV-1991 (Rel. 20, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Sulfate permease II.

GN Cys-14.

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.

OX NCBI_TaxID=5141;

OX NCBI_TaxID=5141;

RP SEQUENCE FROM N.A.

RX MEDLINE=91129256; PubMed=1825178;

RA Ketter J.S., Jaraí G., Pu Y.-H., Marzluf G.A.;

RT "Nucleotide sequence, messenger RNA stability, and DNA recognition elements of cys-14, the structural gene for sulfate permease II in Neurospora crassa."

RL Biochemistry 30:1780-1787(1991).

RL [2]

RP PROBABLE REVISIONS.

RX MEDLINE=94188926; PubMed=8140616;

RA Sandal N.N., Marcker K.A.;

RT "Similarities between a soybean nodulin, Neurospora crassa sulphate permease II and a putative human tumour suppressor.";

RL Trends Biochem. Sci. 19:19-19(1994).

CC -1- FUNCTION: UPTAKE OF SULFATE INTO THE CELL.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- INDUCTION: highly expressed, but only in cells subject to sulfur limitation, and it is turned on by the positive-acting Cys-3 sulfury regulatory protein.

CC -1- MISCELLANEOUS: SULFATE PERMEASE II IS MAINLY FOUND IN MYCELIA.

CC -1- SIMILARITY: BELONGS TO THE SLC26A FAMILY OF TRANSPORTERS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; M59167; AAA33615.1; ALT SEQ.

DR InterPro; IPR001902; Sulph transpt.

DR Pfam; PF00916; Sulfate transp. 1.

DR TIGRfam; TIGR00815; sulP; 1.

DR PROSITE; PS01130; SLC26A; 1.

KW Transport; Transmembrane; Glycoprotein.

FT TRANSMEM 71 91 POTENTIAL.

FT TRANSMEM 103 123 POTENTIAL.

FT TRANSMEM 128 148 POTENTIAL.

FT TRANSMEM 171 191 POTENTIAL.

FT TRANSMEM 193 213 POTENTIAL.

FT TRANSMEM 271 291 POTENTIAL.

FT TRANSMEM 326 346 POTENTIAL.

FT TRANSMEM 363 383 POTENTIAL.

FT TRANSMEM 451 471 POTENTIAL.

FT TRANSMEM 474 494 POTENTIAL.

FT CARBOHYD 23 23

FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).

SO SEQUENCE 788 AA; 87864 MW; 4FC604B60798CE77 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 788;

Best Local Similarity 66.7%; Pred. No. 32;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11

DB 90 VVPQGMAYA 98

RESULT 7

A10C_HUMAN STANDARD; PRT; 1499 AA.

AC O60312; O96914;

DT 30-MAY-2000 (Rel. 39, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Potential phospholipid-transporting ATPase VC (EC 3.6.3.1) (ATPVC)

DE (Amnophospholipid translocase VC).

GN ATP10C OR ATPVC OR KIA0566.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

OX NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.

RX MEDLINE=21225279; PubMed=11326269;

RA Meguro M., Kashiwagi A., Mitsuoka K., Nakao M., Kondo I., Saitoh S.,

RA Oshimura M.;

RT "A novel maternally expressed gene, ATP10C, encodes a putative aminophospholipid translocase associated with Angelman syndrome.";

RL Nat. Genet. 28:19-20(2001).

RL [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=21313119; PubMed=11353404;

RA Herzog L.B.K., Kim S.-J., Cook E.H. Jr., Ledbetter D.H.;

RT "The human aminophospholipid-transporting ATPase gene ATP10C maps

RT adjacent to UBE3A and exhibits similar imprinted expression.";
 RL Am. J. Hum. Genet. 68:1501-1505(2001).
 RN [3]
 RP SEQUENCE OF 337-1499 FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE=98290545; PubMed=9628581;
 RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
 RA Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. IX.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 5:31-39(1998).
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -1- TISSUE SPECIFICITY: Wide expression, with highest levels in
 CC kidney, followed by lung, brain, prostate, testis, ovary, and
 CC small intestine.
 CC -1- DISEASE: Defects in ATP10C are a cause of Angelman syndrome (AS)
 CC [MIM:105830]; also known as 'happy puppet syndrome'. It is
 CC characterized by features of severe motor and intellectual
 CC retardation, microcephaly, ataxia, frequent jerky limb movements
 CC and flapping of the arms and hands, hypotonia, hyperactivity,
 CC hypopigmentation, seizures, absence of speech, frequent smiling
 CC and episodes of paroxysmal laughter, and an unusual facies
 CC characterized by macrostomia, a large mandible and open-mouthed
 CC expression, a great propensity for protruding the tongue ('longue
 CC thrusting'), and an occipital groove.
 CC -1- SIMILARITY: Belongs to the cation transport ATPase family (P-type
 CC ATPases). Subfamily IV.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB051358; BAB47392.1; -;
 DR EMBL; AY029504; AAK33100.1; -;
 DR EMBL; AY029487; AAK33100.1; JOINED.
 DR EMBL; AY029488; AAK33100.1; JOINED.
 DR EMBL; AY029489; AAK33100.1; JOINED.
 DR EMBL; AY029490; AAK33100.1; JOINED.
 DR EMBL; AY029491; AAK33100.1; JOINED.
 DR EMBL; AY029492; AAK33100.1; JOINED.
 DR EMBL; AY029493; AAK33100.1; JOINED.
 DR EMBL; AY029494; AAK33100.1; JOINED.
 DR EMBL; AY029495; AAK33100.1; JOINED.
 DR EMBL; AY029496; AAK33100.1; JOINED.
 DR EMBL; AY029497; AAK33100.1; JOINED.
 DR EMBL; AY029498; AAK33100.1; JOINED.
 DR EMBL; AY029499; AAK33100.1; JOINED.
 DR EMBL; AY029500; AAK33100.1; JOINED.
 DR EMBL; AY029501; AAK33100.1; JOINED.
 DR EMBL; AY029502; AAK33100.1; JOINED.
 DR EMBL; AY029503; AAK33100.1; JOINED.
 DR EMBL; AB011138; BAA25492.1; -;
 DR EMBL; AB011138; BAA25492.1; -;
 DR GeneW; HGNC:13547; ATP10C.
 DR MIM; 605855; -;
 DR MIM; 105830; -;
 DR GO; GO:0016021; C: integral to membrane; NAS.
 DR GO; GO:0004012; F: phospholipid-translocating ATPase activity; NAS.
 DR GO; GO:0008360; F: regulation of cell shape; NAS.
 DR InterPro; IPR001157; ATPase_EI-E2.
 DR InterPro; IPR005834; Filippase.
 DR Pfam; PF00702; Hydrolyase_1.
 DR PRINTS; PR00119; CATRTPASE.
 DR TIGRFAMs; TIGR01652; ATPase-plipid; 1.
 DR TIGRFAMs; TIGR01494; ATPase_P-type; 6.
 DR PROSITE; PS00154; ATPASE_EI_E2; 1.
 KW Hydrolyase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;

KW Multigene family.
 FT DOMAIN 1 86
 FT TRANSMEM 87 106
 FT DOMAIN 107 110
 FT TRANSMEM 111 128
 FT DOMAIN 129 309
 FT TRANSMEM 310 332
 FT DOMAIN 333 362
 FT TRANSMEM 363 384
 FT DOMAIN 385 1087
 FT TRANSMEM 1088 1108
 FT DOMAIN 1109 1119
 FT TRANSMEM 1120 1140
 FT DOMAIN 1141 1170
 FT TRANSMEM 1171 1192
 FT DOMAIN 1193 1199
 FT TRANSMEM 1200 1222
 FT DOMAIN 1223 1228
 FT TRANSMEM 1229 1249
 FT TRANSMEM 1250 1267
 FT TRANSMEM 1268 1292
 FT DOMAIN 1293 1499
 FT MOD_RES 427 427
 FT METAL 1031 1031
 FT METAL 1035 1035
 FT DOMAIN 467 470
 FT CONFLICT 388 388
 SQ SEQUENCE 1499 AA; 167687 MW; D4996A4D0635A68D CRC64;
 Query Match 65.4%; Score 34; DB 1; Length 1499;
 Best Local Similarity 72.7%; Pred. No. 64;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 BEVVPXGMSYS 11
 DB 469 BEVVPFGGSVS 479
 RESULT 8
 RL20_MYCPU STANDARD; PRT; 116 AA.
 AC Q98GV0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 50S ribosomal protein L20.
 GN RPLT OR MYPU 2610.
 OS Mycoplasma pulmonis.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2107;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UAB CTIP;
 RX MEDLINE=21267165; PubMed=11351084;
 RA Chambaud I., Hellig R., Ferris S., Barbe V., Samson D., Galisson F.,
 RA Mozer I., Dybvis K., Wroblewski H., Viart A., Rocha E.P.C.,
 RA Blanchard A.;
 RT "The complete genome sequence of the murine respiratory pathogen
 RT Mycoplasma pulmonis.";
 RL Nucleic Acids Res. 29:2145-2153(2001).
 CC -1- FUNCTION: This protein binds directly to 23s ribosomal RNA and is
 CC necessary for the in vitro assembly process of the 50S ribosomal
 CC subunit. It is not involved in the protein synthesizing functions
 CC of that subunit (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE L20P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

CC -----
DR EMBL; AL445563; CAC13434.1; -.
DR PIR; E90544; E90544.
DR MyPULst; MYPUL_2610; -.
DR HAMAP; MF_00382; -. 1.
DR InterPro; IPR005813; L20.
DR InterPro; IPR005812; L20_bact.org.
DR Pfam; PF00453; Ribosomal_L20_1.
DR PRINTS; PR00062; RIBOSOMAL_L20.
DR PRODOM; PD002389; L20; 1.
DR TIGRFAMs; TIGR01032; rplT_bact; 1.
DR PROSITE; PS00937; RIBOSOMAL_L20; 1.
KM Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 116 AA; 13565 MW; C59C748901B18F14 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 116;
Best Local Similarity 77.8%; Pred. No. 7.1;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 EVVPGMSYS 11
Db 68 VRPLGMSYS 76

RESULT 9
YF49_ARCFU STANDARD; PRT; 165 AA.
AC O28F30;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein AFI949.
GN AFI949.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klank H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrleides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kitzes S.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Macon T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RL Nature 390:364-370 (1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000968; AAB89307.1; -.
DR PIR; D69493; D69493.
DR TIGR; AFI949; -.
KM Hypothetical protein; Transmembrane, Complete proteome.
FT TRANSMEM 7 27 POTENTIAL.
SQ SEQUENCE 141 161 17588 MW; B8C17054810ADB8 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 165;

```

```

Best Local Similarity 60.0%; Pred. No. 10;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVPGMSYS 10
Db 60 BESIDPGASY 69

RESULT 10
Y990_CAMJBE STANDARD; PRT; 253 AA.
ID Y990_CAMJBE
AC P45489; Q9PNV0;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein Cj0990c.
GN Cj0990c.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kelsey J.M., Churcher C.,
RA Baeham D., Chillingworth T., Davies R.M., Felwell T., Holtroyd S.,
RA Jogle K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrall B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668 (2000).
[2]
RN SEQUENCE OF 160-253 FROM N.A.
RC STRAIN=ATCC 43431 / TGH 9011;
RX MEDLINE=95247673; PubMed=7730270;
RA Hart E.K., Chan V.L.;
RT "Expression and characterization of Campylobacter jejuni
RT benzoylglycine amidohydrolase (Hypuricase) gene in Escherichia
RT coli."
RL J. Bacteriol. 177:2396-2402 (1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL139076; CAB73246.1; -.
DR EMBL; Z36940; CA85392.1; -.
DR PIR; C81374; C81374.
DR PIR; I40758; I40758.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 253 AA; 29783 MW; F96D3FF3265F8A6A CRC64;

Query Match 63.5%; Score 33; DB 1; Length 253;
Best Local Similarity 55.6%; Pred. No. 16;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVPGMSYS 10
Db 185 DIPSPGMSY 193

RESULT 11
CTX3_MOUSE STANDARD; PRT; 280 AA.
ID CTX3_MOUSE
AC Q9D387; Q9CXQ4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)

```

DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Protein C20orf103 homolog precursor.
 GN C20ORF103.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISORFMS 1 AND 2).
 RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aizawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Satoh K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saio R.,
 RA Kadota K., Matuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Glasl C., King B., Kochia H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Bonfelli D., Bojunga N., Carninci P., de Bernaldo M.F.,
 RA Brownstein M.O., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guelinich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazarrelli J., Mombarts P.,
 RA Norcone P., Rang B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyshew-Bois A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [2]
 RP SEQUENCE FROM N.A. (ISORFMS 2).
 RC TISSUE=Mammary fibroblast;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Ditchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stjepanovic M., Soares M.B., Bonaldi M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loguelfano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultky S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Hellon E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.T., Skalska U., Smalins D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9D387-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9D387-2; Sequence=VSP_003820;
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
 CC in positions 174 and 239.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use. By non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL, AK014127; BAB29169.1; -.
 DR EMBL, AK018222; BAB31124.1; ALT_FRAME.
 DR EMBL, EC004791; AAH04791.1; -.
 DR MGD; WGI:1920368; 3110035N03RIK.
 DR MGD; WGI:1923411; 6330527006RIK.
 KW Transmembrane; Signal; Alternative splicing.
 FT SIGNAL 1 29
 FT CHAIN 30 280
 FT DOMAIN 30 235
 FT TRANSMEM 236 256
 FT DOMAIN 257 280
 FT CARBOHYD 35 35
 FT CARBOHYD 53 53
 FT CARBOHYD 102 102
 FT CARBOHYD 127 127
 FT VARSPIC 1 118
 FT CONFLICT 221 221
 FT CONFLICT 230 230
 FT CONFLICT 238 238
 SQ SEQUENCE 280 AA; 31721 MW; FA1ID7BF9FDSCCEP CRC64;
 Query Match 63.5%; Score 33; DB 1; Length 280;
 Best Local Similarity 75.0%; Pred. No. 18;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 3 VVPXGMSY 10
 DB 173 VTPAGMSY 180
 RESULT 12
 AROA_VIBCH STANDARD; PRT; 426 AA.
 AC Q9XKB0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
 DE enolpyruvylshikimate-3-phosphate synthase) (EPSPS).
 GN AROA OR VC11732.
 OS Vibrrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=El Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Hatt D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Ueberlack T., Fleischmann R.D., Niernan W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
 RT *cholerae*.";
 RL Nature 406:477-483(2000).
 CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
 CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
 CC PATHWAY: Aromatic amino acid biosynthesis; shikimate pathway;
 CC sixth step.
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- SIMILARITY: Belongs to the EPSP synthase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use. By non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

CC or send an email to license@lsb.ch).
CC -----
CC EMBL, AE004251, AA94882.1; -.
CC DR PIR; D82163; D82163.
CC DR TIGR; VC1732; -.
CC DR HAMAP; MF_00210; -.
CC DR InterPro; IPR006264; AROA.
CC DR InterPro; IPR001986; EPSP synthase.
CC Pfam; PF00275; EPSP_synthase; 1.
CC Prodom; PD001867; EPSP_synthase; 1.
CC DR TIGRFAMs; TIGR01356; AROA; 1.
CC DR PROSITE; PS00104; EPSP SYNTHASE 1; 1.
CC DR PROSITE; PS00885; EPSP SYNTHASE 2; 1.
CC KM Aromatic amino acid biosynthesis; Transferrase; Complete proteome.
CC SEQUENCE 426 AA; 46101 MW; 388526433BFC6; CRC64;

Query Match
Best Local Similarity 60.0%; Pred. No. 28;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
DB 223 EFVIPAQGSY 232

RESULT 13
CC37_SCHPO STANDARD; PRT; 466 AA.
ID CC37_SCHPO
AC 094740;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Hsp90 co-chaperone Cdc37 (Hsp90 chaperone protein kinase-targeting subunit) (Cell division control protein 37).
GN CDC37 OR SPBC9B6.10.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA Westwood P.K., Preston N.C., Fantes P.A.;
RT "Schizosaccharomyces pombe cdc37 gene."
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.

RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gabel K., Goble A., Hamlin N., Harris P., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Nislett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Shenvs K.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymprez B.,
RA Wellens I., Vanstele E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel K., Fuchs M., Fritsch C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Weiler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleau V., Motier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garçon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cernutti L., Lowe T., McCombie J.B., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Useery D., Bartell B.G., Nurse P.;

```

```

RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -1- FUNCTION: With Hsp90 it forms a complex that binds to several
CC kinases, resulting in stabilization and promotion of their
CC activity (By similarity).
CC -1- SUBUNIT: Forms a complex with Hsp90. Interacts with a number of
CC kinases (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE CDC37 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.lsb.ch/announce/
CC or send an email to license@lsb.ch).
CC -----
CC EMBL; AJ132377; CAB38758.1; -.
CC DR EMBL; AJ132376; CAB38757.1; -.
CC DR EMBL; AL049769; CAB42371.2; -.
CC DR PIR; T43653; T43653.
CC DR GenesD_Spombe; SPBC9B6.10; -.
CC DR InterPro; IPR004918; Cdc37.
CC Pfam; PF03234; Cdc37; 1.
CC KM Chaperone; Cell division; Cell cycle.
CC SEQUENCE 466 AA; 52554 MW; 647238B34CAB3C5 CRC64;

Query Match
Best Local Similarity 50.0%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
DB 98 DSAIPGMSY 107

RESULT 14
GSR2_HUMAN STANDARD; PRT; 478 AA.
ID GSR2_HUMAN
AC 09NZM5; Q9B7C6; Q9HAX6; Q9NP1; Q9NPR4; Q9UF12;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glioma tumor suppressor candidate region gene 2 protein (p60).
GN GLTSCR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.
RA MEDLINE=20175430; PubMed=10708517;
RA Smith J.S., Tachibana I., Pohl U., Lee H.K., Thararajasingam U.,
RA Portler B.P., Ueki K., Billings S., Ramaswamy S., Mohrenweiser H.W.,
RA Scheithauer B.W., Louis D.N., Jenkins R.B.;
RT "A transcript map of the chromosome 19q-Arm glioma tumor suppressor
RT region."
RL Genomics 64:44-50(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle; PubMed=12477932;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klauberger R.L., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Datschenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Tomihata S., Cantinici P., Prange C.,
RA Raza S.S., Loguclano N.A., Peters G.J., Abramson J.D., Mullany S.J.,
RA Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,

```

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Paine J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield V.S.N., Krzywinski M.T., Skalska U., Smailus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [3]
 RP SEQUENCE OF 9-478 FROM N.A., AND SUBCELLULAR LOCATION.
 RX MEDLINE=99214318; PubMed=10196275;
 RA Bruni R., Fineschi B., Ogile W.O., Roizman B.,
 RT "A novel cellular protein, p60, interacting with both herpes simplex
 RT virus 1 regulatory proteins ICP22 and ICP0 is modified in a
 RT cell-type-specific manner and is recruited to the nucleus after
 RT infection.";
 RL J. Virol. 73:3810-3817(1999).
 [4]
 RP SEQUENCE OF 12-478 FROM N.A.
 RA Andreu N., Estivill X., Escarceller M., Sunney L.,
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 [5]
 RP SEQUENCE OF 218-477 FROM N.A.
 RC TISSUE=Testis;
 RA Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.,
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: Interacts with HSV-1 early proteins ICP22 and ICP0.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: Expressed at high levels in heart and
 CC pancreas, moderate levels in placenta, liver, skeletal muscle, and
 CC kidney, and low levels in brain and lung.
 CC -1- SIMILARITY: BELONGS TO THE GLTSCR2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF182076; AAF62873.1; -
 DR EMBL; BC004229; AAH04229.1; -
 DR EMBL; BC006311; AAH06311.1; -
 DR EMBL; BC010095; AAH10095.1; -
 DR EMBL; AF296124; AAG30413.1; -
 DR EMBL; AL359335; CAB94786.1; -
 DR EMBL; AL359336; CAB94787.1; -
 DR EMBL; AL122063; CAB59242.1; -
 DR SWISS-2DPAGE: Q9WZM5; HUMAN.
 DR Genew; HGNC:4333; GLTSCR2.
 DR MIM; 605691; -
 DR GO; GO:0008181; F: tumor suppressor; TAS.
 KW Nuclear protein; Polymorphism.
 FT VARIANT 389 389
 FT R -> Q.
 FT /FTID=VAR_011486.
 FT GGS -> HEQ (IN REF. 2; AAH04229).
 FT G -> R (IN REF. 3).
 FT RRKEQLMEKLAKOGELPREVRAQAQLNPSATRAKPGPD
 FT TVERP -> SGRSSYGRMPSPRASSPGAGQSPVPAQPCFN
 FT KGPNPAPGHRITAA (IN REF. 3).
 FT SDNPLRDLNCGODEFFLE -> LNNPDPVVPWPGCLFPG
 FT (IN REF. 3).
 FT A -> S (IN REF. 2; AAH04229).
 FT D -> H (IN REF. 3).
 FT PEGNLTDRDFKSFGRNMIEPRERAKFKKTKVKLVKRAAF
 FT REIQ -> VLVYSCRGACPMVTEPLLVPRGRGRHHGCP
 FT WAGPYGMPRG (IN REF. 5).
 FT EGNILDRFKSFGRNMIEPRERAKFKKTKVKLVKRAAF
 FT EIQL -> RQGHSEFTGRARRGGI (IN REF. 3).
 FT SEQUENCE 478 AA; 54417 MM; 7F18923E348CB52B CRC64;

Query Match 63.5%; Score 33; DB 1; Length 478;
 Best Local Similarity 60.0%; Pred. No. 31;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 EVVPRKMSYS 11
 DB 239 EVAPAGASYN 248
 RESULT 15
 BCN5_CLOPE STANDARD; PRT; 890 AA.
 ID BCN5_CLOPE
 AC P08696;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE Bacteriocin BCN5.
 GN BCN.
 OS Clostridium perfringens.
 OG Plasmid pIP404.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 CC Clostridium.
 CX NCBI_TaxID=1502;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CPN50;
 RX MEDLINE=88336297; PubMed=2901768;
 RA Garnier T., Cole S.T.,
 RT "Complete nucleotide sequence and genetic organization of the
 RT bacteriocinogenic plasmid, pIP404, from Clostridium perfringens.";
 RL Plasmid 19:114-150(1988).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CPN50.
 RX MEDLINE=87057020; PubMed=2877971;
 RA Garnier T., Cole S.T.,
 RT "Characterization of a bacteriocinogenic plasmid from Clostridium
 RT perfringens and molecular genetic analysis of the
 RT bacteriocin-encoding gene.";
 RL J. Bacteriol. 168:1189-1196(1986).
 [3]
 RP SEQUENCE OF 1-14 FROM N.A.
 RC STRAIN=CPN50;
 RX MEDLINE=89039249; PubMed=2460717;
 RA Garnier T., Cole S.T.,
 RT "Studies of UV-inducible promoters from Clostridium perfringens in
 RT vivo and in vitro.";
 RL Mol. Microbiol. 2:607-614(1988).
 CC -1- FUNCTION: MAY FUNCTION AS AN IONOPHORE.
 CC -1- INDUCTION: By UV irradiation.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M14481; AAA98248.1; -
 DR EMBL; M32882; AAA98249.1; -
 DR PIR; A30481; A30481.
 DR InterPro; IPR003646; SH3_bac.
 DR InterPro; IPR00834; Zn_carboxypept.
 DR Pfam; PF00246; Zn_carboxypept; 1.
 DR SMART; SM00287; SH3b; 3.
 KW Antibiotic; Bacteriocin; Plasmid.
 FT DOMAIN 815 869 HYDROPHOBIC.
 FT SEQUENCE 890 AA; 96699 MM; F4B5E8971C316C6 CRC64;
 Query Match 63.5%; Score 33; DB 1; Length 890;
 Best Local Similarity 66.7%; Pred. No. 60;

Matches	6;	Conservative	1;	Mismatches	2;	Indels	0;	Gaps	0;
Qy	2	EVPXGMSY	10						
			:						
Db	170	EVPGGFTY	178						

Search completed: December 22, 2003, 17:42:23
Job time : 4.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using BW model

Run on: December 22, 2003, 17:27:26 ; Search time 25.2 Seconds

(without alignments)
112.642 Million cell updates/sec

Title: US-09-909-164-6

Perfect score: 52
Sequence: 1 EEVVPXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacterioph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	75.0	1044	16	Q8DIH0 synechococ
2	38	73.1	387	16	Q98FX1 rhizobium 1
3	38	73.1	3472	1	Q74056 cenarchaeum
4	37	71.2	840	3	Q9URY8 echinosacch
5	37	71.2	1123	16	Q8EMD4 mus musculu
6	36	69.2	471	11	Q8R126 mus musculu
7	36	69.2	484	11	Q8VDI8 mus musculu
8	36	69.2	484	11	Q8BTK4 mus musculu
9	36	69.2	484	11	Q8BK35 mus musculu
10	35	67.3	225	10	Q40129 lycopersico
11	35	67.3	425	5	Q9XVK4 caenorhabdi
12	35	67.3	495	11	Q8C1D7 mus musculu
13	35	67.3	556	4	Q43733 homo sapien
14	35	67.3	583	5	Q9BH83 plasmodium
15	35	67.3	583	5	Q9BH85 plasmodium
16	35	67.3	583	5	Q81587 plasmodium

17	35	67.3	670	11	Q01487 ratus ratt
18	35	67.3	747	16	Q8PM16 xanthomonas
19	35	67.3	747	16	Q8PAT2 xanthomonas
20	35	67.3	749	16	Q9PDM6 xylella fas
21	35	67.3	1902	4	Q14122 homo sapien
22	34	65.4	156	3	Q12479 saccharomyc
23	34	65.4	219	17	Q971S2 sulfolobus
24	34	65.4	252	17	Q28342 archaeoglob
25	34	65.4	290	4	Q96MT1 homo sapien
26	34	65.4	387	16	Q92MD6 rhizobium m
27	34	65.4	489	4	Q81YMS homo sapien
28	34	65.4	541	16	Q98BP5 ureaplasma
29	34	65.4	544	16	Q9PQD2 ureaplasma
30	34	65.4	842	3	Q9URR4 penicillium
31	34	65.4	899	16	Q8G4I5 bifidobacte
32	34	65.4	1049	16	Q8XT05 ralbionia 8
33	33	63.5	143	17	Q8TX62 methanopyru
34	33	63.5	166	16	Q8PPP5 xanthomonas
35	33	63.5	193	2	Q8VUA8 lactococcus
36	33	63.5	208	2	Q8KTO4 candidatus
37	33	63.5	209	16	Q8RRE5 fusobacteri
38	33	63.5	284	16	P74187 synechocyst
39	33	63.5	298	10	Q9M3C0 arabidopsis
40	33	63.5	326	12	Q9QCE7 soil-borne
41	33	63.5	326	12	Q9Q9Q7 soil-borne
42	33	63.5	326	12	Q9DJG4 soil-borne
43	33	63.5	326	12	Q9Q9Q9 soil-borne
44	33	63.5	326	12	Q91DN1 soil-borne
45	33	63.5	326	12	Q9Q9Q5 soil-borne

ALIGNMENTS

RESULT 1	Q8DIH0	PRELIMINARY;	PRT; 1044 AA.
AC	Q8DIH0:		
DT	01-MAR-2003 (TREMBLrel_23, Created)		
DT	01-MAR-2003 (TREMBLrel_23, Last sequence update)		
DT	01-MAR-2003 (TREMBLrel_23, Last annotation update)		
DE	Multidrug efflux transporter.		
GN	TLL1618.		
OS	Synechococcus elongatus (Thermosynechococcus elongatus).		
OC	Bacteria; Cyanobacteria; Chroococcales; Synechococcus.		
OX	NCBI_TaxID=32046;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BP-1;		
RX	MEDLINE=2225144; PubMed=12240834;		
RA	Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,		
RA	Watanabe A., Iriyuchi M., Kawashima K., Kimura T., Kishida Y.,		
RA	Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,		
RA	Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;		
RT	"Complete genome structure of the thermophilic cyanobacterium		
RT	Thermosynechococcus elongatus BP-1.";		
RL	DNA Rep. 9:123-130(2002).		
DR	EMBL; AP005374; BAC09170.1; -.		
KW	Complete proteome.		
SQ	SEQUENCE 1044 AA; 113205 MW; 00E9C13F0F636D2F CRC64;		
Query Match	75.0%; Score 39; DB 16; Length 1044;		
Beat Local Similarity	63.6%; Pred. No. 24;		
Matches	7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;		
QY	1 EEVVPXGMSYS 11		
DB	843 EEVVPXGMSYS 853		
RESULT 2	Q98FX1	PRELIMINARY;	PRT; 387 AA.
ID	Q98FX1		

AC 098FX1;
 DT 01-OCT-2001 (TReMBLrel. 18, Created)
 DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
 DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
 DE Hippurate hydrolase.
 GN MLR3583.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF03099;
 RX MEDLINE=21082930; PubMed=1121968;
 RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idehara K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpoto S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti."
 RL DNA Res. 7:331-338 (2000).
 DR EMBL; AP003002; BAB5045.1; -
 DR InterPro: IPR002933; Peptidase_M20.
 DR Pfam; PF01546; Peptidase_M20; I.
 KM Hydrolyase; Complete proteome.
 SQ SEQUENCE 387 AA; 41180 MW; 13BFF8E64306829 CRC64;
 QY Query Match 73.1%; Score 38; DB 16; Length 387;
 Best Local Similarity 60.0%; Pred. No. 13;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Db 367 DEAIPIHGMSY 376
 QY 1 EEVVPXGMSY 10
 ID 074056 PRELIMINARY; PRT; 3472 AA.
 AC 074056;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Hypothetical 367.1 kDa protein.
 OS Cenarchaeum symbiosum.
 OC Archaea; Crenarchaeota; Thermoprotei; Cenarchaeales; Cenarchaeaceae;
 OC Cenarchaeum.
 OX NCBI_TaxID=46770;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B;
 RX MEDLINE=98422450; PubMed=9748430;
 RA Schleper C., DeLong E.F., Preston C.M., Feldman R.A., Wu K.Y.,
 RA Swanson R.V.;
 RT "Genomic analysis reveals chromosomal variation in natural populations
 RT of the uncultured psychrophilic archaeon Cenarchaeum symbiosum."
 RL J. Bacteriol. 180:5003-5009 (1998).
 DR EMBL; AF083072; AAC62699.1; -
 DR InterPro: IPR000515; BPD transp.
 DR InterPro: IPR001680; WD40.
 DR Pfam; PF00400; WD40; 4.
 DR SMART; SM00320; WD40; 2.
 DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBER; 1.
 KM Hypothetical protein; Repeat; WD repeat.
 SQ SEQUENCE 3472 AA; 367058 MW; 37F80707030F9355 CRC64;
 QY Query Match 73.1%; Score 38; DB 1; Length 3472;
 Best Local Similarity 54.5%; Pred. No. 1.5e+02;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 Db 1 EEVVPXGMSY 11
 ID 074056 PRELIMINARY; PRT; 3472 AA.
 AC 074056;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Hypothetical 367.1 kDa protein.
 OS Cenarchaeum symbiosum.
 OC Archaea; Crenarchaeota; Thermoprotei; Cenarchaeales; Cenarchaeaceae;
 OC Cenarchaeum.
 OX NCBI_TaxID=46770;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B;
 RX MEDLINE=98422450; PubMed=9748430;
 RA Schleper C., DeLong E.F., Preston C.M., Feldman R.A., Wu K.Y.,
 RA Swanson R.V.;
 RT "Genomic analysis reveals chromosomal variation in natural populations
 RT of the uncultured psychrophilic archaeon Cenarchaeum symbiosum."
 RL J. Bacteriol. 180:5003-5009 (1998).
 DR EMBL; AF083072; AAC62699.1; -
 DR InterPro: IPR000515; BPD transp.
 DR InterPro: IPR001680; WD40.
 DR Pfam; PF00400; WD40; 4.
 DR SMART; SM00320; WD40; 2.
 DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBER; 1.
 KM Hypothetical protein; Repeat; WD repeat.
 SQ SEQUENCE 3472 AA; 367058 MW; 37F80707030F9355 CRC64;

Db 2294 EDVIRGISFS 2304
 QY Query Match 71.2%; Score 37; DB 3; Length 840;
 Best Local Similarity 77.8%; Pred. No. 51;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 135 VVPGMSYA 143
 QY 3 VVPGMSYS 11
 ID 08BMD4 PRELIMINARY; PRT; 1123 AA.
 AC 08BMD4;
 DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE MYPE 2560 paralog, 57%.
 GN MYPE2710.
 OS Mycoplasma penetrans.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=28227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HE-2;
 RX MEDLINE=22354719; PubMed=12466555;
 RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
 RA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
 RT "The complete genomic sequence of Mycoplasma penetrans, an
 RT intracellular bacterial pathogen in humans."
 RL Nucleic Acids Res. 30:5293-5300 (2002).
 DR EMBL; AP004171; BAC44062.1; -
 DR NCBI; AP004171; BAC44062.1; -
 KM Complete proteome.
 SQ SEQUENCE 1123 AA; 123636 MW; AAD707330E3DB4AC CRC64;
 QY Query Match 71.2%; Score 37; DB 16; Length 1123;
 Best Local Similarity 70.0%; Pred. No. 70;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Db 2 EVVVPXGMSY 11
 ID 08BMD4 PRELIMINARY; PRT; 1123 AA.
 AC 08BMD4;
 DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE MYPE 2560 paralog, 57%.
 GN MYPE2710.
 OS Mycoplasma penetrans.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=28227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HE-2;
 RX MEDLINE=22354719; PubMed=12466555;
 RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
 RA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
 RT "The complete genomic sequence of Mycoplasma penetrans, an
 RT intracellular bacterial pathogen in humans."
 RL Nucleic Acids Res. 30:5293-5300 (2002).
 DR EMBL; AP004171; BAC44062.1; -
 DR NCBI; AP004171; BAC44062.1; -
 KM Complete proteome.
 SQ SEQUENCE 1123 AA; 123636 MW; AAD707330E3DB4AC CRC64;

RESULT 6
 Q8R126 PRELIMINARY; PRT; 471 AA.
 AC Q8R126; 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Hypothetical 54.5 kDa protein (Fragment).
 GN GLTSCR2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Strausberg R.;
 RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; BC025810; AA025810.1; --
 DR MGI; MGI:2154441; Gltsctr2.
 KW Hypothetical protein.
 FT NON TER
 SQ SEQUENCE 471 AA; 54506 MW; EDDA685C374A9760 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 471;
 Best Local Similarity 60.0%; Pred. No. 44;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVFXGMSYS 11
 DB 226 EVIPAGASYN 235

RESULT 7

Q8VD18 PRELIMINARY; PRT; 484 AA.
 AC Q8VD18; 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Similar to glioma tumor suppressor candidate region gene 2.
 GN GLTSCR2 OR AMS36441.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Salivary gland;
 RA Strausberg R.;
 RL Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; BC017637; AA017637.1; --
 DR MGI; MGI:2154441; Gltsctr2.
 SQ SEQUENCE 484 AA; 55835 MW; BBB45F3B4BE02A36 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
 Best Local Similarity 60.0%; Pred. No. 45;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVFXGMSYS 11
 DB 239 EVIPAGASYN 248

RESULT 8

Q8BTX4 PRELIMINARY; PRT; 484 AA.
 AC Q8BTX4; 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE Similar to glioma tumor suppressor CANDIDATE region gene 2
 DE Protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NOD; TISSUE=Thymus;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK088461; BAC40367.1; --
 SQ SEQUENCE 484 AA; 55806 MW; B3056425B5EBCA08 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
 Best Local Similarity 60.0%; Pred. No. 45;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVFXGMSYS 11
 DB 239 EVIPAGASYN 248

RESULT 9

Q8BK35 PRELIMINARY; PRT; 484 AA.
 AC Q8BK35; 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Similar to glioma tumor suppressor CANDIDATE region gene 2
 DE Protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Plutitary;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK077341; BAC36760.1; --
 SQ SEQUENCE 484 AA; 55792 MW; BB67949BCB92D44 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
 Best Local Similarity 60.0%; Pred. No. 45;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVFXGMSYS 11
 DB 239 EVIPAGASYN 248

RESULT 10

Q40129 PRELIMINARY; PRT; 225 AA.
 AC Q40129; 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Hypothetical 25.2 kDa protein precursor.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Jamids; Solanales; Solanales; Solanales; Solanaceae; Solanum.

OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VP36; TISSUE=Plac1;
RX MEDLINE=95375233; PubMed=7647701;
RA Milligan S.B., Gasser C.S.;
RT "Nature and regulation of plac1-expressed genes in tomato."
RL Plant Mol. Biol. 28:691-711(1995).
DR EMBL, U20592; AAA80497.1;
DR InterPro; IPR002160; Kunitz_legume.
DR Pfam; PF00197; Kunitz_legume; 1.
DR PRINTS; PR00291; KUNITZINBTR.
DR ProDom; PD000891; Kunitz_legume; 1.
DR SMART; SM00452; ST1; 1.
DR PROSITE; PS00283; SOYBEAN_KUNITZ; 1.
KM Hypothetical protein; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 225 UNKNOWN.
SQ SEQUENCE 225 AA; 25188 MW; 1074C261D20CFDAD CRC64;

Query Match 67.3%; Score 35; DB 10; Length 225;
Best Local Similarity 54.5%; Pred. No. 32;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 11
Db 32 DEVVPMGKTYA 42

RESULT 11

O9XVK4 PRELIMINARY; PRT; 425 AA.
ID O9XVK4
AC O9XVK4
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE R10D12.10 protein.
GN R10D12.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Percy C.M.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; Z81109; CAB03241.1;
DR WormPep; R10D12.10; CE12690.
DR InterPro; IPR000719; Prot_kinase.
DR Pfam; PF00069; kinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KM ATP-binding; Transferase.
SQ SEQUENCE 425 AA; 49410 MW; 5D96E29B08C8E9D6 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 425;
Best Local Similarity 50.0%; Pred. No. 64;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 EEVVPXGMSY 10
Db 335 EGVVPGGLQY 344

RESULT 12

O8CID7

ID O8CID7 PRELIMINARY; PRT; 495 AA.
AC O8CID7;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE CDNA FLJ31891 FIS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the Riken Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
Nature 420:563-573(2002).
RL EMBL; AK028258; BAC25846.1;
DR EMBL; AK028258; BAC25846.1;
SQ SEQUENCE 495 AA; 55358 MW; B6A22A093C114752 CRC64;

Query Match 67.3%; Score 35; DB 11; Length 495;
Best Local Similarity 63.6%; Pred. No. 75;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 11
Db 260 EEIVPMGISPS 270

RESULT 13

O43733 PRELIMINARY; PRT; 556 AA.
ID O43733
AC O43733;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-JUN-1998 (TREMBlrel. 06, Last annotation update)
DE DNA binding protein (Fragment).
GN D451B15.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tuddy B.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z98050; CAB10847.1;
FT NON TER 1
SQ SEQUENCE 556 AA; 59059 MW; ECB00E4033FB2528 CRC64;

Query Match 67.3%; Score 35; DB 4; Length 556;
Best Local Similarity 66.7%; Pred. No. 86;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 VVPXGMSYS 11
Db 244 VVPAGLTYS 252

RESULT 14

O9BH83 PRELIMINARY; PRT; 583 AA.
ID O9BH83
AC O9BH83;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Choline transporter.
GN SCTL.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Ben Mamoun C., Gluzman I.Y., Goldberg D.E.;
 RT "Plasmodium falciparum choline transporter (PfSCT1) gene."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY007374; AAK14818.1; -
 DR EMBL; AY007373; AAK14817.1; -
 DR InterPro: IPR002123; Acyltransferase.
 DR Pfam: PF01553; Acyltransferase; 1.
 DR SMART; SM00563; PlSC; 1.
 SQ SEQUENCE 583 AA; 66918 MW; 4DF83D7530527474 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 583;
 Best Local Similarity 55.6%; Pred. No. 90;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
 :||:||||
 Db 227 IIPVGLSYS 235

RESULT 15
 Q9BHA5 PRELIMINARY; PRT; 583 AA.
 ID Q9BHA5
 AC Q9BHA5;
 DT 01-JUN-2001 (T-EMBLrel. 17, Created)
 DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE Choline transporter.
 GN SCT1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxId=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ben Mamoun C., Gluzman I.Y., Goldberg D.E.;
 RT "Plasmodium falciparum choline transporter (PfSCT1) gene."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY007372; AAK14816.1; -
 DR EMBL; AY007375; AAG17947.1; -
 DR InterPro: IPR002123; Acyltransferase.
 DR Pfam: PF01553; Acyltransferase; 1.
 DR SMART; SM00563; PlSC; 1.
 SQ SEQUENCE 583 AA; 66917 MW; 2B2BFAE3E395E049 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 583;
 Best Local Similarity 55.6%; Pred. No. 90;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
 :||:||||
 Db 227 IIPVGLSYS 235

Search completed: December 22, 2003, 17:51:24
 Job time : 26.2667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:41:00 ; Search time 32.4667 Seconds
(without alignments)
53.776 Million cell updates/sec

Title: US-09-909-164-7

Perfect score: 56

Sequence: 1 EEVVPXGMHYS 11

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	54	96.4	11 23	ABR80523
2	54	96.4	11 23	ABR80527
3	54	96.4	11 23	ABR80558
4	54	96.4	11 23	ABR80537
5	49	87.5	11 23	ABR80537
6	49	87.5	11 23	ABR80541
7	48	85.7	11 23	ABR80546
8	48	85.7	11 23	ABR80550
9	48	85.7	11 23	ABR80554

10	48	85.7	11 23	ABR80555	Hepatitis C virus
11	46	82.1	11 23	ABR80531	Hepatitis C virus
12	46	82.1	11 23	ABR80532	Hepatitis C virus
13	45	80.4	11 23	ABR80521	Hepatitis C virus
14	45	80.4	11 23	ABR80522	Hepatitis C virus
15	45	80.4	11 23	ABR80524	Hepatitis C virus
16	45	80.4	11 23	ABR80525	Hepatitis C virus
17	45	80.4	11 23	ABR80526	Hepatitis C virus
18	45	80.4	11 23	ABR80528	Hepatitis C virus
19	45	80.4	11 23	ABR80529	Hepatitis C virus
20	45	80.4	11 23	ABR80559	Hepatitis C virus
21	45	80.4	11 23	ABR80561	Hepatitis C virus
22	45	80.4	11 23	ABR80562	Hepatitis C virus
23	45	80.4	11 23	ABR80563	Hepatitis C virus
24	45	80.4	11 23	ABR80564	Hepatitis C virus
25	45	80.4	11 23	ABR80565	Hepatitis C virus
26	45	80.4	11 23	ABR80566	Hepatitis C virus
27	45	80.4	11 23	ABR80567	Hepatitis C virus
28	45	80.4	11 23	ABR80568	Hepatitis C virus
29	40	71.4	11 23	ABR80535	Hepatitis C virus
30	40	71.4	11 23	ABR80536	Hepatitis C virus
31	40	71.4	11 23	ABR80538	Hepatitis C virus
32	40	71.4	11 23	ABR80539	Hepatitis C virus
33	40	71.4	11 23	ABR80540	Hepatitis C virus
34	40	71.4	11 23	ABR80542	Hepatitis C virus
35	40	71.4	11 23	ABR80543	Hepatitis C virus
36	39	69.6	11 23	ABR80544	Hepatitis C virus
37	39	69.6	11 23	ABR80545	Hepatitis C virus
38	39	69.6	11 23	ABR80547	Hepatitis C virus
39	39	69.6	11 23	ABR80548	Hepatitis C virus
40	39	69.6	11 23	ABR80549	Hepatitis C virus
41	39	69.6	11 23	ABR80551	Hepatitis C virus
42	39	69.6	11 23	ABR80552	Hepatitis C virus
43	39	69.6	11 23	ABR80553	Hepatitis C virus
44	39	69.6	11 23	ABR80556	Hepatitis C virus
45	39	69.6	11 23	ABR80557	Hepatitis C virus

ALIGNMENTS

RESULT 1
ABR80523
ID ABR80523 standard; peptide; 11 AA.
AC ABR80523;
XX
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #3.
XX
KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketamide;
vtrncide.
XX
OS Synthetic.
XX
FH
FT Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT FT
FT Modified-site 6 /note= "N-terminal acetyl"
FT FT
FT FT
FT Misc-difference 9 /note= "D-form residue"
FT FT
FT Modified-site 11 /note= "C-terminal amide"
FT FT
PM WO200208251-A2.
PD 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX

PR 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
XX
XX Llm-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus.
XX
XX Sequence 11 AA;
SQ

Query Match 96.4%; Score 54; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00053; Indels 0; Gaps 0;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 EEVVPXGMHYS 11
Db 1 EEVVPXGMHYS 11

RESULT 2
ABB80527
ID ABB80527 standard; peptide; 11 AA.
XX
XX ABB80527;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #7.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 1 /note= "N-terminal acetyl"
XX FT 6
XX Modified-site /note= "N-terminyl carbonyl forming keto-amide linkage with
XX FT residue 7"
XX FT 8
XX Misc-difference /note= "D-form residue"
XX FT 11
XX Modified-site /note= "C-terminal amide"
XX FT
XX W0200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Llm-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX

PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus.
XX
XX Sequence 11 AA;
SQ

Query Match 96.4%; Score 54; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00053;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 EEVVPXGMHYS 11
Db 1 EEVVPXGMHYS 11

RESULT 3
ABB80558
ID ABB80558 standard; peptide; 11 AA.
XX
XX ABB80558;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #38.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 1 /note= "N-terminal acetyl"
XX FT 6
XX Modified-site /note= "N-terminyl carbonyl forming keto-amide linkage with
XX FT residue 7"
XX FT 8
XX Modified-site /note= "Oxymethionine"
XX FT 11
XX Modified-site /note= "C-terminal amide"
XX FT
XX W0200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Llm-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease -
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 96.4%; Score 54; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00053;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BEVVPXGMHYS 11
|||
1 BEVVPXGMHYS 11

DB 1 BEVVPXGMHYS 11

RESULT 4
ABB80560
ID ABB80560 standard; peptide; 11 AA.

XX ABB80560;

DT 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #40.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM virucide.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "N-terminal acetyl" forming keto-amide linkage with
FT residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Modified-site 8 /note= "Oxymethionine"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

FT W0200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Lin-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory

XX activity useful for treating disorders associated with hepatitis C

XX Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient

CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 96.4%; Score 54; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00053;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BEVVPXGMHYS 11
|||
1 BEVVPXGMHYS 11

DB 1 BEVVPXGMHYS 11

RESULT 5
ABB80537
ID ABB80537 standard; peptide; 11 AA.

XX ABB80537;

DT 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #17.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM virucide.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "N-terminal acetyl" forming keto-amide linkage with
FT residue 7"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

FT W0200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Lin-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory

XX activity useful for treating disorders associated with hepatitis C

XX Claim 17; Page 64; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 87.5%; Score 49; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0051;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMHYS 11
 |||||
 DB 1 EEVVPXGMHYS 11

RESULT 6

ABB80541 standard; peptide; 11 AA.

08-OCT-2002 (first entry)

Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #21.

Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketamide;
 virucide.

Synthetic.

Key Location/Qualifiers

Modified-site 1 /note= "N-terminal acetyl"

Modified-site 6 /note= "N-terminal carbonyl forming keto-amide linkage with
 residue 7"

Misc-difference 8 /note= "D-form residue"

Modified-site 11 /note= "C-terminal amide"

MO200208251-A2.

31-JAN-2002.

19-JUL-2001; 2001WO-US23169.

21-JUL-2000; 2000US-220101P.

(CORV-) CORVAS INT INC.

Lim-wilby M, Levy OE, Brunck TK;

WPI; 2002-361643/39.

Novel peptide compound having hepatitis C virus protease inhibitory

activity useful for treating disorders associated with hepatitis C

virus protease -

Claim 17; Page 65; 69pp; English.

The sequence represents a peptide compound of the invention having
 hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 invention are alpha-ketamide peptide analogues. The peptides have
 virucide activity, and are useful for treating and in the manufacture of
 a medicament to treat disorders associated with HCV protease. A
 pharmaceutical composition comprising the peptide as an active ingredient
 is useful for treating disorders associated with hepatitis C virus.

Sequence 11 AA;

Query Match 87.5%; Score 49; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.0051;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMHYS 11
 |||||
 DB 1 EEVVPXGMHYS 11

RESULT 7
 ABB80546
 ID ABB80546 standard; peptide; 11 AA.

XX ABB80546;
 AC
 XX
 DT 08-OCT-2002 (first entry)

Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #26.

Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketamide;
 virucide.

Synthetic.

Key Location/Qualifiers

Modified-site 1 /note= "N-terminal acetyl"

Modified-site 6 /note= "N-terminal carbonyl forming keto-amide linkage with
 residue 7"

Modified-site 11 /note= "C-terminal amide"

MO200208251-A2.

31-JAN-2002.

19-JUL-2001; 2001WO-US23169.

21-JUL-2000; 2000US-220101P.

(CORV-) CORVAS INT INC.

Lim-wilby M, Levy OE, Brunck TK;

WPI; 2002-361643/39.

Novel peptide compound having hepatitis C virus protease inhibitory

activity useful for treating disorders associated with hepatitis C

virus protease -

Claim 17; Page 65; 69pp; English.

The sequence represents a peptide compound of the invention having
 hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 invention are alpha-ketamide peptide analogues. The peptides have
 virucide activity, and are useful for treating and in the manufacture of
 a medicament to treat disorders associated with HCV protease. A
 pharmaceutical composition comprising the peptide as an active ingredient
 is useful for treating disorders associated with hepatitis C virus.

Sequence 11 AA;

Query Match 85.7%; Score 48; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.008;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMHYS 11
 |||||
 DB 1 EEVVPXGMHYS 11

RESULT 8

ABB80550 standard; peptide; 11 AA.

08-OCT-2002 (first entry)

Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #30.

Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketamide;
 virucide.

OS Synthetic.
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6.
FT /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
FT
PN WO200208251-A2.
XX
XX 31-JAN-2002.
PD
XX 19-JUL-2001; 2001WO-US23169.
PF
XX 21-JUL-2000; 2000US-220101P.
PR
XX (CORV-) CORVAS INT INC.
PA
XX Lim-wilby M, Levy OE, Brunck TK;
PI
XX WPI; 2002-361643/39.
DR
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 65; 69pp; English.
PS
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
CC
XX
SQ Sequence 11 AA;
Query Match 85.7%; Score 48; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.008; 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGSHYS 11
Db 1 EEVVPXGSHYS 11
RESULT 9
ABB80554 ID ABB80554 standard; peptide; 11 AA.
XX
XX ABB80554;
AC
XX
XX 08-OCT-2002 (first entry)
DT
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #34.
DE
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM virucide.
XX
XX Synthetic.
OS
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT

FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
FT
PN WO200208251-A2.
XX
XX 31-JAN-2002.
PD
XX 19-JUL-2001; 2001WO-US23169.
PF
XX 21-JUL-2000; 2000US-220101P.
PR
XX (CORV-) CORVAS INT INC.
PA
XX Lim-wilby M, Levy OE, Brunck TK;
PI
XX WPI; 2002-361643/39.
DR
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 65; 69pp; English.
PS
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
CC
XX
SQ Sequence 11 AA;
Query Match 85.7%; Score 48; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.008; 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGSHYS 11
Db 1 EEVVPXGSHYS 11
RESULT 10
ABB80555 ID ABB80555 standard; peptide; 11 AA.
XX
XX ABB80555;
AC
XX
XX 08-OCT-2002 (first entry)
DT
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #35.
DE
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM virucide.
XX
XX Synthetic.
OS
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
XX
PN WO200208251-A2.

XX 31-JAN-2002.
PD 19-JUL-2001; 2001WO-US23169.
XX 21-JUL-2000; 2000US-220101P.
PR (CORV-) CORVAS INT INC.
XX Lim-wilby M, Levy OE, Brunck TK;
PI WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
XX virus protease -
XX Claim 17; Page 65; 69pp; English.
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX Sequence 11 AA;
SQ
Query Match 85.7%; Score 48; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.008; Mismatches 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMHYS 11
DB 1 EEVVPXGSHYS 11
RESULT 11
ABB80531 ID ABB80531 standard; peptide; 11 AA.
XX ABB80531;
AC 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #11.
DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketamide;
KM virucide.
XX Synthetic.
OS Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11 /note= "C-terminal amide"
FT WO200208251-A2.
PN 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US23169.
PF 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
PA Lim-wilby M, Levy OE, Brunck TK;
PI

XX WPI; 2002-361643/39.
DR Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX Claim 17; Page 64; 69pp; English.
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX Sequence 11 AA;
SQ
Query Match 82.1%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.02; Mismatches 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMHYS 11
DB 1 EEVVPXGSHYS 11
RESULT 12
ABB80532 ID ABB80532 standard; peptide; 11 AA.
XX ABB80532;
AC 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #12.
DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketamide;
KM virucide.
XX Synthetic.
OS Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT WO200208251-A2.
PN 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US23169.
PF 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
PA Lim-wilby M, Levy OE, Brunck TK;
PI WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
PS Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX
SQ Sequence 11 AA;

Query Match 82.1%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMYHS 11
|||
DB 1 EEVVPXGMYHS 11

RESULT 13

ABB80521 standard; peptide; 11 AA.

AC ABB80521;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
virocide.

OS Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"

FT Modified-site 11 /note= "C-terminal amide"

PN WO200208251-A2.

PD 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

PA (CORV-) CORVAS INT INC.

PI Lim-wilby M, Levy OE, Brunck TK;

DR WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -

PS Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX
SQ Sequence 11 AA;

Query Match 80.4%; Score 45; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.031;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMYHS 11
|||
DB 1 EEVVPXGMYHS 11

RESULT 14

ABB80522 standard; peptide; 11 AA.

AC ABB80522;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
virocide.

OS Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

PN WO200208251-A2.

PD 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

PA (CORV-) CORVAS INT INC.

PI Lim-wilby M, Levy OE, Brunck TK;

DR WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -

PS Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX
SQ Sequence 11 AA;

Query Match 80.4%; Score 45; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.031;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMYHS 11
|||
DB 1 EEVVPXGMYHS 11

RESULT 15

ID ABB80524 standard; peptide; 11 AA.

XX ABB80524;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.

KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.

OS Synthetic.

XX Key Location/Qualifiers

FH Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

FT

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease -

XX Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus.

SQ Sequence 11 AA;

Query Match 80.4%; Score 45; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.031; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMHYS 11
 |||||
 Db 1 EEVVPXGMDYS 11

Search completed: December 22, 2003, 17:41:00
 Job time : 32.4667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: December 22, 2003, 16:37:03 ; Search time 10.1333 Seconds
(without alignment)
45.930 Million cell updates/sec

Title: US-09-909-164-7
Perfect score: 56
Sequence: 1 EEVVPXGMHYS 11

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	69.6	1037	4	US-09-134-001C-4794
2	37	66.1	856	4	US-09-252-991A-21444
3	34	60.7	600	2	US-08-821-119-19
4	34	60.7	600	2	US-08-821-118-2
5	33	58.9	277	4	US-09-252-991A-26615
6	33	58.9	385	4	US-09-252-991A-27834
7	33	58.9	747	4	US-09-724-864-36
8	32	57.1	70	4	US-09-134-001C-3950
9	32	57.1	102	2	US-08-580-988A-23
10	32	57.1	126	2	US-08-879-995A-3
11	32	57.1	126	3	US-09-215-096-3
12	32	57.1	152	3	US-08-460-694-4
13	32	57.1	152	3	US-08-460-744-4
14	32	57.1	152	3	US-07-667-711B-4
15	32	57.1	173	1	US-08-193-977-7
16	32	57.1	189	2	US-08-464-517-21
17	32	57.1	189	2	US-08-246-361A-21
18	32	57.1	189	3	US-08-463-772-21
19	32	57.1	189	5	PCT-US93-05000-21
20	32	57.1	236	2	US-08-464-517-22
21	32	57.1	236	2	US-08-246-361A-22
22	32	57.1	236	3	US-08-463-772-22
23	32	57.1	236	5	PCT-US93-05000-22
24	32	57.1	280	2	US-08-464-517-6
25	32	57.1	280	3	US-08-463-772-6
26	32	57.1	289	2	US-08-246-361A-4
27	32	57.1	289	5	PCT-US93-05000-4

28	32	57.1	291	5	PCT-US93-05000-6	Sequence 6, Appli
29	32	57.1	292	2	US-08-464-517-23	Sequence 23, Appli
30	32	57.1	292	2	US-08-246-361A-6	Sequence 6, Appli
31	32	57.1	292	2	US-08-246-361A-23	Sequence 23, Appli
32	32	57.1	292	3	US-08-463-772-23	Sequence 23, Appli
33	32	57.1	292	5	PCT-US93-05000-23	Sequence 23, Appli
34	32	57.1	295	1	US-07-947-120-8	Sequence 8, Appli
35	32	57.1	295	2	US-08-472-893A-8	Sequence 8, Appli
36	32	57.1	295	1	US-08-460-694-2	Sequence 2, Appli
37	32	57.1	295	2	US-08-464-517-19	Sequence 19, Appli
38	32	57.1	295	2	US-08-464-517-20	Sequence 20, Appli
39	32	57.1	295	2	US-08-246-361A-19	Sequence 19, Appli
40	32	57.1	295	2	US-08-246-361A-20	Sequence 20, Appli
41	32	57.1	295	3	US-08-463-772-19	Sequence 19, Appli
42	32	57.1	295	3	US-08-463-772-20	Sequence 20, Appli
43	32	57.1	295	3	US-08-460-744-2	Sequence 2, Appli
44	32	57.1	295	3	US-07-667-711B-2	Sequence 2, Appli
45	32	57.1	295	3	US-08-947-492-8	Sequence 8, Appli

ALIGNMENTS

```
RESULT 1
US-09-134-001C-4794
Sequence 4794, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GNC-007
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4794
LENGTH: 1037
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4794

Query Match      69.6% Score 39; DB 4; Length 1037;
Best Local Similarity 63.6% Pred. NO. 20;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Cy 1 EEVVPXGMHYS 11
Db 199 KEVSNGLHYS 209

RESULT 2
US-09-252-991A-21444
Sequence 21444, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 21444
LENGTH: 856
TYPE: PRT
```

ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21444

Query Match 66.1%; Score 37; DB 4; Length 856;

Best Local Similarity 70.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMHY 10
|||
64 EAVVGGSHY 73

RESULT 3

US-08-821-119-19

Sequence 19, Application US/08821119

Patent No. 5821104

GENERAL INFORMATION:

APPLICANT: Holm, Kaj Andre

APPLICANT: Rasmussen, Grethe

APPLICANT: Halckier, Torben

APPLICANT: Lembeck, Jan

TITLE OF INVENTION: Tripeptidyl Aminopeptidase

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESSES:

ADDRESSEE: No. 58211040 No. 5821104disk of No. 5821104th America, Inc.

STREET: 405 Lexington Avenue

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10174

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/821.119

FILING DATE: 19-MAR-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J

REGISTRATION NUMBER: 33,728

REFERENCE/DOCKET NUMBER: 4107.204-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655

TELEX:

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 600 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

US-08-821-119-19

Query Match 60.7%; Score 34; DB 2; Length 600;

Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 VPKGMHYS 11
|||
31 VPKGMHYS 38

RESULT 4

US-08-821-118-2

Sequence 2, Application US/08821118

Patent No. 5989889

GENERAL INFORMATION:

APPLICANT: Rey, Michael

APPLICANT: Golightly, Elizabeth

TITLE OF INVENTION: NUCLEIC ACIDS ENCODING POLYPEPTIDES

TITLE OF INVENTION: HAVING TRIPEPTIDE AMINOPEPTIDASE

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESSES:

ADDRESSEE: No. 59898890 No. 5989889disk of No. 5989889th America, Inc.

STREET: 405 Lexington Avenue

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10174

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/821.118

FILING DATE: 19-MAR-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J

REGISTRATION NUMBER: 33,728

REFERENCE/DOCKET NUMBER: 4107.400-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 600 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

US-08-821-118-2

Query Match 60.7%; Score 34; DB 2; Length 600;

Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 VPKGMHYS 11
|||
31 VPKGMHYS 38

RESULT 5

US-09-252-991A-26615

Sequence 26615, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: ABRIGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252.991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074.788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094.190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 31142

SEQ ID NO 26615

TYPE: PRT

LENGTH: 277

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-26615

Query Match 58.9%; Score 33; DB 4; Length 277;

Best Local Similarity 63.6%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 EEVFXGMHYS 11
|||
Db 48 EETVFGGHTS 58

RESULT 6
US-09-252-991A-27834
; Sequence 27834, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107136.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27834
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27834

Query Match 58.9%; Score 33; DB 4; Length 385;
Best Local Similarity 44.4%; Pred. No. 1e+02;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 EEVFXGMHYS 10
|||
Db 201 EILPRLALHY 209

RESULT 7
US-09-724-864-36
; Sequence 36, Application US/09724864
; Patent No. 6380362
; GENERAL INFORMATION:
; APPLICANT: Watson, James D
; APPLICANT: Murison, James G.
; TITLE OF INVENTION: Polynucleotides, polypeptides expressed
; TITLE OF INVENTION: by the polynucleotides and methods for their use.
; FILE REFERENCE: 11000.1050U
; CURRENT APPLICATION NUMBER: US/09/724,864
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: U.S. No. 6380362 60/171,678
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 747
; TYPE: PRT
; ORGANISM: Rat
US-09-724-864-36

Query Match 58.9%; Score 33; DB 4; Length 747;
Best Local Similarity 71.4%; Pred. No. 2.1e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 5 PXGMHYS 11
|||
Db 627 PGGLHYS 633

RESULT 8
US-09-134-001C-3950
; Sequence 3950, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCI
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3950
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3950

Query Match 57.1%; Score 32; DB 4; Length 70;
Best Local Similarity 62.5%; Pred. No. 24;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 4 VPKGMHYS 11
:|:|:|
Db 36 WPKGFHYS 43

RESULT 9
US-08-580-988A-23
; Sequence 23, Application US/08580988A
; Patent No. 5856161
; GENERAL INFORMATION:
; APPLICANT: Aggarwal et al.
; TITLE OF INVENTION: Tumor Necrosis Factor
; TITLE OF INVENTION: Receptor-I-Associated Protein Kinase And Methods
; TITLE OF INVENTION: For Its Use
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. Benjamin A. Adler
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 Mb floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/580,988A
; FILING DATE: January 3, 1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benjamin Aaron Adler, Ph.D., J.D.
; REGISTRATION NUMBER: 35,423
; REFERENCE/DOCKET NUMBER: D5721CIP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-777-2321
; TELEFAX: 713-777-6908
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 102 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: no
; HYPOTHEICAL: no
; ANTI-SENSE: no
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:

US-08-580-988A-23

Query Match 57.1%; Score 32; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 37;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMH 10
DB 24 EEVFPAMNY 33

RESULT 10

US-08-879-995A-3
; Sequence 3, Application US/08879995A
; Patent No. 5985606
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: HUMAN PREPROTACHYKININ B
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/879,995A
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0326 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 163590
; US-08-879-995A-3

Query Match 57.1%; Score 32; DB 2; Length 126;
Best Local Similarity 66.7%; Pred. No. 46;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMH 9
DB 28 EQVVPGGGH 36

RESULT 11
US-09-215-096-3
; Sequence 3, Application US/09215096
; Patent No. 6008194
; GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Kaser, Matthew R.

TITLE OF INVENTION: HUMAN PREPROTACHYKININ B

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/215,096

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/879,995

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0326 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX:

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 126 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 163590

US-09-215-096-3

Query Match 57.1%; Score 32; DB 3; Length 126;
Best Local Similarity 66.7%; Pred. No. 46;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMH 9
DB 28 EQVVPGGGH 36

RESULT 12

US-08-460-694-4
; Sequence 4, Application US/08460694
; Patent No. 5858655
; GENERAL INFORMATION:
; APPLICANT: Arnold, Andrew
; TITLE OF INVENTION: PRAD1 Cyclin and its cDNA
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/460,694
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McConathy, Evelyn H.
REGISTRATION NUMBER: 35,279
REFERENCE/DOCKET NUMBER: 0609.4070002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 152 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-460-694-4

Query Match 57.1%; Score 32; DB 2; Length 152;
Best Local Similarity 60.0%; Pred. No. 57;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EEVVPXGMHY 10
Db 20 EEVFPPLANNY 29

RESULT 13
US-08-460-744-4
Sequence 4, Application US/08460744
Patent No. 6107541
GENERAL INFORMATION:
APPLICANT: Arnold, Andrew
TITLE OF INVENTION: Prad1 Cyclin and its cDNA
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,744
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McConathy, Evelyn H.
REGISTRATION NUMBER: 35,279
REFERENCE/DOCKET NUMBER: 0609.4070005
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 152 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-460-744-4

Query Match 57.1%; Score 32; DB 3; Length 152;
Best Local Similarity 60.0%; Pred. No. 57;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Qy 1 EEVVPXGMHY 10

Db 20 EEVFPPLANNY 29

RESULT 14
US-07-667-711B-4
Sequence 4, Application US/07667711B
Patent No. 6110700
GENERAL INFORMATION:
APPLICANT: ARNOLD, ANDREW
TITLE OF INVENTION: Prad1 Cyclin and its cDNA
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVE., NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/667,711B
FILING DATE: 11-MAR-1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MCPHAIL, DONALD R.
REGISTRATION NUMBER: 35,811
REFERENCE/DOCKET NUMBER: 0609.4070000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 152 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
US-07-667-711B-4

Query Match 57.1%; Score 32; DB 3; Length 152;
Best Local Similarity 60.0%; Pred. No. 57;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EEVVPXGMHY 10
Db 20 EEVFPPLANNY 29

RESULT 15
US-08-193-977-7
Sequence 7, Application US/08193977
Patent No. 5625031
GENERAL INFORMATION:
APPLICANT: WEBSTER, KEVIN R.
APPLICANT: COLEMAN, KEVIN G.
TITLE OF INVENTION: PEPTIDE INHIBITORS OF THE P33CDK2 AND
TITLE OF INVENTION: P33CDK2 CELL CYCLE REGULATORY KINASES AND HUMAN
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: REED & ROBINS
STREET: 635 BRYANT STREET
CITY: PAJO ALTO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/193,977
FILING DATE: 08-FEB-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: ROBINS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 5998-0016
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 617-8999
TELEFAX: (415) 327-3231
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 173 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-193-977-7

Query Match 57.1%; Score 32; DB 1; Length 173;
Best Local Similarity 60.0%; Pred. No 65;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 EEVVPXGMHY 10
||| |:
Db 55 EEVFPPLANNY 64

Search completed: December 22, 2003, 16:43:41
Job time : 11.1333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 16:40:14 / Search time 20.8667 Seconds
(without alignments)
98.451 Million cell updates/sec

Title: US-09-909-164-7

Perfect score: 56

Sequence: 1 EEVVPXGMHYS 11

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	58.9	273	US-10-094-749-2319	Sequence 2319, App
2	33	58.9	567	US-10-270-333-126	Sequence 126, App
3	33	58.9	622	US-09-738-626-4919	Sequence 4919, App
4	33	58.9	747	US-09-866-050A-663	Sequence 663, App
5	33	58.9	3472	US-10-029-120-4	Sequence 4, App1
6	33	58.9	3472	US-10-027-806-4	Sequence 4, App1
7	33	58.9	3472	US-10-034-623-4	Sequence 4, App1
8	33	58.9	3472	US-10-027-801-4	Sequence 4, App1
9	32	57.1	36	US-10-199-820-241	Sequence 241, App
10	32	57.1	236	US-10-029-386-32076	Sequence 32076, App
11	32	57.1	242	US-10-094-749-2076	Sequence 2076, App
12	32	57.1	254	US-09-778-927A-53	Sequence 53, App1
13	32	57.1	276	US-10-116-275-114	Sequence 114, App
14	32	57.1	289	US-09-919-497-54	Sequence 54, App1
15	32	57.1	289	US-10-024-066-2	Sequence 2, App1

16	32	57.1	289	US-10-024-066-4	Sequence 4, App1
17	32	57.1	292	US-10-116-275-275	Sequence 275, App
18	32	57.1	315	US-10-238-075-1315	Sequence 1315, App
19	32	57.1	381	US-09-769-738-24	Sequence 24, App1
20	32	57.1	519	US-09-934-455-164	Sequence 164, App
21	32	57.1	519	US-10-302-267-36	Sequence 36, App1
22	32	57.1	653	US-09-820-843A-26	Sequence 26, App1
23	32	57.1	715	US-09-252-088-16	Sequence 16, App1
24	32	57.1	715	US-10-340-792-16	Sequence 16, App1
25	32	57.1	793	US-09-252-088-15	Sequence 15, App1
26	32	57.1	793	US-10-340-792-15	Sequence 15, App1
27	32	57.1	822	US-09-769-736-18	Sequence 18, App1
28	31.5	56.2	662	US-10-047-542-80	Sequence 80, App1
29	31.5	56.2	847	US-09-870-759-52	Sequence 52, App1
30	31.5	56.2	847	US-09-751-708A-52	Sequence 52, App1
31	31.5	56.2	847	US-10-207-653-133	Sequence 133, App
32	31	55.4	59	US-09-948-080-14	Sequence 14, App1
33	31	55.4	73	US-09-864-761-40832	Sequence 40832, App
34	31	55.4	153	US-10-029-386-32486	Sequence 32486, App
35	31	55.4	192	US-09-986-480-171	Sequence 171, App
36	31	55.4	192	US-10-001-857-119	Sequence 119, App
37	31	55.4	216	US-10-156-761-12762	Sequence 12762, App
38	31	55.4	251	US-10-106-698-6257	Sequence 6257, App
39	31	55.4	260	US-09-815-242-13489	Sequence 13489, App
40	31	55.4	260	US-09-815-242-13613	Sequence 13613, App
41	31	55.4	299	US-09-815-242-10697	Sequence 10697, App
42	31	55.4	336	US-09-782-974C-86	Sequence 86, App1
43	31	55.4	337	US-09-943-198-4	Sequence 4, App1
44	31	55.4	337	US-09-885-453-1	Sequence 1, App1
45	31	55.4	337	US-10-017-161-526	Sequence 526, App

ALIGNMENTS

RESULT 1
US-10-094-749-2319, Application US/10094749
Sequence 2319, US20030219741A1
Publication No. US20030219741A1
GENERAL INFORMATION:
APPLICANT: ISOGAI, TAKAO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: OTSUKI, TETSUJI
APPLICANT: OTSUKI, TETSUJI
APPLICANT: WAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHITO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTOTYUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
FILE REFERENCE: 084335/0160
CURRENT APPLICATION NUMBER: US/10/094, 749
PRIOR FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350, 435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 381
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 2319
LENGTH: 273
TYPE: PRT
ORGANISM: Homo sapiens
US-10-094-749-2319

Query Match 58.9%; Score 33; DB 12; Length 273;
Best Local Similarity 71.4%; Pred. No. 1.7e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 PEXGMHY 11
|:|:|:|
DB 151 PGGLHY 157

RESULT 2
US-10-270-333-126

Sequence 126, Application US/10270333
Publication No. US20030092124A1

GENERAL INFORMATION:

APPLICANT: Cravchik, Anibal

TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS, AND USES

FILE REFERENCE: CL000733CON

CURRENT APPLICATION NUMBER: US/10/270,333

PRIOR FILING DATE: 2002-10-15

PRIOR APPLICATION NUMBER: 60/168,677

PRIOR FILING DATE: 1999-12-03

PRIOR APPLICATION NUMBER: 60/175,691

PRIOR FILING DATE: 2000-01-12

PRIOR APPLICATION NUMBER: 60/191,638

PRIOR FILING DATE: 2000-03-23

NUMBER OF SEQ ID NOS: 198

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 126

LENGTH: 567

TYPE: PRT

ORGANISM: Drosophila

US-10-270-333-126

Query Match 58.9%; Score 33; DB 15; Length 567;
Best Local Similarity 83.3%; Pred. No. 3.7e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 PEXGMHY 10
|:|:|:|
DB 402 PEXGMHY 407

RESULT 3
US-09-738-626-4919

Sequence 4919, Application US/09738626

Publication No. US20020197605A1

GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI

APPLICANT: MIZOGUCHI, HIROSHI

APPLICANT: ANDO, SEIKO

APPLICANT: HAYASHI, MIKIRO

APPLICANT: OCHIAI, KEIKO

APPLICANT: YOKOI, HARUHIKO

APPLICANT: TATEISHI, NAKO

APPLICANT: SENOH, AKIHIRO

APPLICANT: IKEDA, MASATO

APPLICANT: OZAKI, AKIO

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-125

CURRENT APPLICATION NUMBER: US/09/738,626

PRIOR FILING DATE: 2000-12-18

PRIOR APPLICATION NUMBER: JP 99/377484

PRIOR FILING DATE: 1999-12-16

PRIOR APPLICATION NUMBER: JP 00/159162

PRIOR FILING DATE: 2000-04-07

PRIOR APPLICATION NUMBER: JP 00/280988

PRIOR FILING DATE: 2000-08-03

NUMBER OF SEQ ID NOS: 7059

SOFTWARE: PatentIn ver. 3.0

SEQ ID NO 4919

LENGTH: 622
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-4919

Query Match 58.9%; Score 33; DB 10; Length 622;
Best Local Similarity 50.0%; Pred. No. 4.1e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 2; Gaps 1;

QY 1 EEVPEXGM--HY 10
|:|:|:|:|:|
DB 46 EEIPIGVPMHY 57

RESULT 4
US-09-866-050A-663

Sequence 663, Application US/09866050A

Publication No. US20030040471A1

GENERAL INFORMATION:

APPLICANT: Watson, James D.

APPLICANT: Strachan, Lorna

APPLICANT: Sleeman, Matthew

APPLICANT: Onrust, Rene

APPLICANT: Murison, James G.

APPLICANT: Kumble, Krishanand D.

TITLE OF INVENTION: Compositions Isolated From Skin Cells

FILE REFERENCE: 11000.1011c4U

CURRENT APPLICATION NUMBER: US/09/866,050A

PRIOR FILING DATE: 2001-05-24

NUMBER OF SEQ ID NOS: 725

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 663

LENGTH: 747

TYPE: PRT

ORGANISM: Rat

US-09-866-050A-663

Query Match 58.9%; Score 33; DB 11; Length 747;
Best Local Similarity 71.4%; Pred. No. 5e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 PEXGMHY 11
|:|:|:|
DB 627 PGGLHY 633

RESULT 5
US-10-029-120-4

Sequence 4, Application US/10029120

Publication No. US20030175708A1

GENERAL INFORMATION:

APPLICANT: Swanson, Ronald V.

APPLICANT: Feldman, Robert A.

APPLICANT: Schleper, Christa

TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM

FILE REFERENCE: DCOB.002A

CURRENT APPLICATION NUMBER: US/10/029,120

PRIOR FILING DATE: 2001-12-21

PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020

PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29

NUMBER OF SEQ ID NOS: 123

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 4

LENGTH: 3472

TYPE: PRT

ORGANISM: Cenarchaeum symbiosum

US-10-029-120-4

Query Match 58.9%; Score 33; DB 12; Length 3472;
Best Local Similarity 45.5%; Pred. No. 2.6e+03;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMHYS 11
|:|:|:|:|:
Db 2294 EDVIRPGISFS 2304

RESULT 6
US-10-027-806-4
; Sequence 4, Application US/10027806
; Publication No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schlieper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOIP.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-806-4

Query Match 58.9%; Score 33; DB 14; Length 3472;
Best Local Similarity 45.5%; Pred. No. 2.6e+03;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMHYS 11
|:|:|:|:|:
Db 2294 EDVIRPGISFS 2304

RESULT 7
US-10-034-623-4
; Sequence 4, Application US/10034623
; Publication No. US20020198365A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schlieper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOIP.002A
; CURRENT APPLICATION NUMBER: US/10/034,623
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-034-623-4

Query Match 58.9%; Score 33; DB 14; Length 3472;
Best Local Similarity 45.5%; Pred. No. 2.6e+03;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMHYS 11
|:|:|:|:|:
Db 2294 EDVIRPGISFS 2304

RESULT 8
US-10-027-801-4
; Sequence 4, Application US/10027801
; Publication No. US20030054364A1

; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schlieper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOIP.002A
; CURRENT APPLICATION NUMBER: US/10/027,801
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-801-4

Query Match 58.9%; Score 33; DB 15; Length 3472;
Best Local Similarity 45.5%; Pred. No. 2.6e+03;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMHYS 11
|:|:|:|:|:
Db 2294 EDVIRPGISFS 2304

RESULT 9
US-10-199-820-241
; Sequence 241, Application US/10199820
; Publication No. US20030180739A1
; GENERAL INFORMATION:
; APPLICANT: Board of Trustees of the University of Illinois
; APPLICANT: Primiano, Thomas
; APPLICANT: Chang, Bey-dih
; APPLICANT: Roninson, Igor
; TITLE OF INVENTION: Methods and Reagents for Identifying Gene Targets for Treating
; FILE REFERENCE: 99,216-U
; CURRENT APPLICATION NUMBER: US/10/199,820
; CURRENT FILING DATE: 2002-09-23
; NUMBER OF SEQ ID NOS: 314
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 241
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-199-820-241

Query Match 57.1%; Score 32; DB 12; Length 36;
Best Local Similarity 60.0%; Pred. No. 31;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMHYS 10
|:|:|:|:|:
Db 20 EEVFPPLANVY 29

RESULT 10
US-10-029-386-32076
; Sequence 32076, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32076

```
LENGTH: 236
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO 298050.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.3
OTHER INFORMATION: SWISSPROT HIT: P15822, EVALUATE 1.00e-125
US-10-029-386-32076

Query Match          57.1%; Score 32; DB 12; Length 236;
Best Local Similarity 66.7%; Pred. No. 2.3e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 3 EVVPGMHYS 11
DB 80 VVPAGLTYS 88

RESULT 11
US-10-094-749-2076
Sequence 2076, Application US/10094749
Publication No. US20030219741A1
GENERAL INFORMATION:
APPLICANT: ISOGAI, TAKAO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: OTSUKI, TETSUJI
APPLICANT: WAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOFARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAORIKO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTOMYUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
FILE REFERENCE: 084335/0160
CURRENT APPLICATION NUMBER: US/10/094,749
CURRENT FILING DATE: 2002-03-12
PRIORITY APPLICATION NUMBER: 60/350,435
PRIORITY FILING DATE: 2002-01-24
PRIORITY APPLICATION NUMBER: JP 2001-328381
PRIORITY FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 3381
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2076
LENGTH: 242
TYPE: PRT
ORGANISM: Homo sapiens
US-10-094-749-2076

Query Match          57.1%; Score 32; DB 12; Length 242;
Best Local Similarity 60.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 EEVVPXGMHY 10
DB 25 EEVPLAMNY 34

RESULT 12
```

```
US-09-778-927A-53
Sequence 53, Application US/09778927A
Patent No. US20020068342A1
GENERAL INFORMATION:
APPLICANT: KHOSRAVI, Rami et al.
TITLE OF INVENTION: NOVEL NUCLEIC ACID AND AMINO ACID SEQUENCES AND NOVEL
TITLE OF INVENTION: VARIANTS OF ALTERNATIVE SPLICING
FILE REFERENCE: 2786-0160P
CURRENT APPLICATION NUMBER: US/09/778,927A
CURRENT FILING DATE: 2001-02-08
PRIORITY APPLICATION NUMBER: IL 134453
PRIORITY FILING DATE: 2000-02-09
PRIORITY APPLICATION NUMBER: IL135341
PRIORITY FILING DATE: 2000-03-29
NUMBER OF SEQ ID NOS: 81
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 53
LENGTH: 254
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)..(254)
OTHER INFORMATION: Xaa = any amino acid, unknown, or other
US-09-778-927A-53

Query Match          57.1%; Score 32; DB 9; Length 254;
Best Local Similarity 60.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 EEVVPXGMHY 10
DB 74 EEVPLAMNY 83

RESULT 13
US-10-116-275-114
Sequence 114, Application US/10116275
Publication No. US20030211476A1
GENERAL INFORMATION:
APPLICANT: Elan Pharmaceutical Technology
APPLICANT: O'Mahony, Daniel J.
APPLICANT: Brayden, David
APPLICANT: Byrne, Deragh
APPLICANT: Lambkin, Imelda
APPLICANT: Higgins, Lisa
TITLE OF INVENTION: Genetic Analysis of Peyer's Patches and M Cells and Methods and
TITLE OF INVENTION: Compositions Targeting Peyer's Patches and M Cell Receptors
FILE REFERENCE: E1067/20087
CURRENT APPLICATION NUMBER: US/10/116,275
CURRENT FILING DATE: 2002-10-04
NUMBER OF SEQ ID NOS: 349
SOFTWARE: PatentIn version 3.1
SEQ ID NO 114
LENGTH: 276
TYPE: PRT
ORGANISM: Homo sapiens
US-10-116-275-114

Query Match          57.1%; Score 32; DB 12; Length 276;
Best Local Similarity 60.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 EEVVPXGMHY 10
DB 75 EEVPLAMNY 84

RESULT 14
US-09-919-497-54
Sequence 54, Application US/09919497
Patent No. US2002010662A1
GENERAL INFORMATION:
```

```

; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 54
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-919-497-54

```

```

Query Match      57.1%; Score 32; DB 10; Length 289;
Best Local Similarity 60.0%; Pred. No. 2.9e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      1 EEVVPXGMHY 10
      ||| | | |
Db      74 EEVFPPLAMNY 83

```

```

RESULT 15
US-10-024-066-2
; Sequence 2, Application US/10024066
; Publication No. US20020166134A1
; GENERAL INFORMATION:
; APPLICANT: Field, Loren J.
; APPLICANT: Pasmamarchi, Kishore Babu S.
; TITLE OF INVENTION: CARDIOMYOCYTES WITH ENHANCED PROLIFERATIVE POTENTIAL,
; TITLE OF INVENTION: AND METHODS FOR PREPARING AND USING SAME
; FILE REFERENCE: 7037-450
; CURRENT APPLICATION NUMBER: US/10/024,066
; CURRENT FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 60/139,942
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: PCT/US00/16827
; PRIOR FILING DATE: 2000-06-19
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Mus musculus
; US-10-024-066-2

```

```

Query Match      57.1%; Score 32; DB 14; Length 289;
Best Local Similarity 60.0%; Pred. No. 2.9e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      1 EEVVPXGMHY 10
      ||| | | |
Db      73 EEVFPPLAMNY 82

```

Search completed: December 22, 2003, 17:32:41
 Job time : 21.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:24:36 / Search time 9.06667 Seconds

(without alignment)
116.675 Million cell updates/sec

Title: US-09-909-164-7
Sequence: 1 EEVVPXGMHY 11

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	66.1	308	2	A72207
2	37	66.1	1057	2	F89892
3	36	64.3	102	2	A42452
4	36	64.3	252	2	AE2001
5	36	64.3	460	2	SE6906
6	36	64.3	743	2	S38143
7	35	62.5	156	2	D82618
8	35	62.5	233	2	T02590
9	35	62.5	311	2	H69194
10	35	62.5	425	2	T24111
11	35	62.5	425	2	G86430
12	34	60.7	264	2	G69117
13	34	60.7	279	2	G75538
14	34	60.7	350	2	B75478
15	34	60.7	355	2	T35025
16	34	60.7	360	2	B69086
17	34	60.7	425	2	C83903
18	34	60.7	426	2	S58132
19	34	60.7	495	2	T28717
20	34	60.7	1028	2	AF3286
21	33	58.9	156	2	S54619
22	33	58.9	367	2	B83607
23	33	58.9	441	2	G82253
24	33	58.9	466	2	G71542
25	33	58.9	466	2	H81697
26	33	58.9	487	2	S65811
27	33	58.9	514	1	HODVLB
28	33	58.9	534	2	A69284
29	33	58.9	545	2	T08564

ALIGNMENTS

30	33	58.9	627	2	A69663	DNA mismatch repair
31	33	58.9	716	1	UC5061	macrophage-stimula
32	33	58.9	1257	2	S44754	CI4B9.8 protein -
33	33	58.9	1396	2	S36851	L-shaped tail fibe
34	33	58.9	3472	2	T31308	hypothetical 367K
35	33	58.9	126	2	A25905	tachylin B precu
36	32	57.1	197	2	D71640	heme exporter prote
37	32	57.1	225	2	S57810	hypothetical prote
38	32	57.1	233	2	B97120	ribosomal protein
39	32	57.1	267	2	T07215	probable Thua prot
40	32	57.1	270	2	C95881	hypothetical prote
41	32	57.1	283	2	T25737	cyclin D2 - rat
42	32	57.1	288	2	UC4011	cyclin D2 - mouse
43	32	57.1	288	2	158372	cyclin D2 - human
44	32	57.1	289	2	A41984	
45	32	57.1	289	2	A42822	

RESULT 1

A72207 fish proteinase activity modulator Hfik - Thermotoga maritima (strain MSB8)

C/Species: Thermotoga maritima

C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C/Accession: A72207

R/Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hich Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, L.C.M.

Nature 399, 323-329, 1999

A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A/Reference number: A72200; MUID:99287316; PMID:10360571

A/Accession: A72207

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-308 <ARN>

A/Cross-references: GB:AE001819; GB:AE000512; NID:94982396; PIDN:ADJ36885.1; PID:94982

A/Experimental source: strain MSB8

A/Genetics:

A/Gene: TM1822

C/Superfamily: erythrocyte band 7 integral membrane protein

Query Match	66.1%	Score 37;	DB 2;	Length 308;
Best Local Similarity	75.0%	Pred. No. 10;		
Matches	6;	Conservative	1;	Mismatches
			1;	Indels
				Gaps
				0;

RESULT 2

F89892 carbamoyl-phosphate synthase large chain [imported] - Staphylococcus aureus (strain N3

C/Species: Staphylococcus aureus

C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C/Accession: F89892

R/Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Og

ma, A.; Mitani, H.; Hattori, M.; Ogasawara, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.

C./Shiba, T.; Hattori, M.; Ogasawara, N.; Hattori, K.

Lancet 357, 1225-1240, 2001

A/Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A/Reference number: A89758; MUID:21311952; PMID:11418146

A/Accession: F89892

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1057 <RUR>

A/Cross-references: GB:BA000018; PID:913701002; PIDN:BA42298.1; GSPDB:GN00149

A/Experimental source: strain N315

A/Genetics:

A/Gene: pyrAB

C/Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain; bioti

Query Match 66.1%; Score 37; DB 2; Length 1057;
 Best Local Similarity 60.0%; Pred. No. 39;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 EVVPXGMHYS 11
 ||| |||
 Db 190 EIVSNGLHYS 199

RESULT 3

VI protein - tobacco yellow dwarf virus (strain Australia)
 C:Species: tobacco yellow dwarf virus
 C:Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
 C:Accession: A42452
 R:Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
 Virology 187, 633-642, 1992
 A:Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow
 A:Reference number: A42452; MUID:92188538; PMID:1546458
 A:Accession: A42452
 A:Molecule type: DNA
 A:Residues: 1-102 <MOR>
 A:Cross-references: GB:M81103; NID:G335283; PIDN:AAA47947.1; PID:G335284

Query Match 64.3%; Score 36; DB 2; Length 102;
 Best Local Similarity 60.0%; Pred. No. 5;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMHYS 11
 ||| |||
 Db 7 QVVPSGINYS 16

RESULT 4

hypothetical protein alr1563 [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AE2001
 R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Iriyuchi, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AE2001
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-252 <KIR>
 A:Cross-references: GB:BA000019; PIDN:BA077929.1; PID:G17135383; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: alr1563

Query Match 64.3%; Score 36; DB 2; Length 252;
 Best Local Similarity 50.0%; Pred. No. 13;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMHYS 10
 ||| |||
 Db 235 EMIVPAGLHF 244

RESULT 5

hypothetical protein YPL139C - yeast (Saccharomyces cerevisiae)
 C:Species: Saccharomyces cerevisiae
 C:Date: 22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 17-Mar-2000
 C:Accession: S69046
 R:Hall, J.; DePaulo, T.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.;
 submitted to the EMBL Data Library, December 1995
 A:Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.

A:Reference number: S69040

A:Accession: S69046
 A:Molecule type: DNA

A:Residues: 1-460 <HAL>

A:Cross-references: EMBL:U43703; NID:G1244769; PIDN:AAB68221.1; PID:G1244776; MIPS:YPL

C:Genetics:

A:Gene: SGD:UME1

A:Cross-references: SGD:S0006060; MIPS:YPL139C

A:Map position: 16L

C:Superfamily: Saccharomyces cerevisiae transcription modulator WTM1

Query Match 64.3%; Score 36; DB 2; Length 460;
 Best Local Similarity 62.5%; Pred. No. 25;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMHYS 10
 ||| |||
 Db 85 IVPLGLHYS 92

RESULT 6

hypothetical protein YBL011W homolog YKR067W - Yeast (Saccharomyces cerevisiae)
 C:Species: Saccharomyces cerevisiae
 C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 19-Apr-2002
 C:Accession: S38143
 R:Van Vlier-Reedijk, J.C.; Planta, R.J.
 submitted to the Protein Sequence Database, March 1994
 A:Reference number: S38130
 A:Accession: S38143
 A:Molecule type: DNA

A:Residues: 1-743 <VAN>

A:Cross-references: EMBL:Z28292; NID:G486536; PIDN:CAA82146.1; PID:G486537; MIPS:YKR06

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:GP72

A:Cross-references: SGD:S0001775

A:Map position: 11R

C:Keywords: transmembrane protein

Query Match 64.3%; Score 36; DB 2; Length 743;
 Best Local Similarity 75.0%; Pred. No. 43;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMHYS 10
 ||| |||
 Db 294 VVPXGLHYS 301

RESULT 7

conserved hypothetical protein XPI950 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: D82618
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequ

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: D82618

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-156 <SIM>

A:Cross-references: GB:AE004014; GB:AE003849; NID:G9107044; PIDN:AAF84752.1; GSPDB:GN0

A:Experimental source: strain 9a5c

R:Simpon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvaranga, R.;

Briener, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre,

as-Melo, B.; Docena, C.; El-Dotry, H.; Facincani, A.F.; Ferreira, A.J.S.

submitted to Genbank, June 2000

A:Authors: Ferreira, V.C.A.; Perito, J.A.; Praga, J.S.; Franca, S.C.; Franco, M.C.; Fro

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.B.; Lai

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins,

A;Authors: Martine, E.M.F.; Mateukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E. de Sa, R.G.; Santelli, R.V.; Sawasak
 A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Teshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A;Reference number: A59328
 A;Contents: annotation
 C;Genetics:
 A;Gene: XPI950

Query Match 62.5%; Score 35; DB 2; Length 156;
 Best Local Similarity 55.6%; Pred. No. 13;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 EEVVPXGMH 9
 |||||
 Db 119 EEIIPGCVH 127

RESULT 8
 T02590
 DNA binding protein EREBP-2 - common tobacco
 C;Species: Nicotiana tabacum (common tobacco)
 C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 21-Jul-2000
 C;Accession: T02590
 R;Ome-Takagi, M.; Shinshi, H.
 Plant Cell 7, 173-182, 1995
 A;Title: Ethylene-inducible DNA binding proteins that interact with an ethylene responsi
 A;Reference number: 214671; PMID:95276459; PMID:7756828
 A;Accession: T02590
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-233 <OHM>
 A;Cross-references: EMBL:D38126; NID:G790362; PIDN:BAA07324.1; PID:G1208498
 A;Experimental source: strain BY4; tissue-type leaf

Query Match 62.5%; Score 35; DB 2; Length 233;
 Best Local Similarity 60.0%; Pred. No. 19;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 EEVVPXGMH 10
 :|||
 Db 90 QAVVPKGRHY 99

RESULT 9
 H69194
 GMP synthetase, subunit B - Methanobacterium thermoautotrophicum (strain Delta H)
 C;Species: Methanobacterium thermoautotrophicum
 C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
 C;Accession: H69194
 R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
 D.; Qiu, D.; Spadafora, R.; Vicafora, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
 ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
 A;Reference number: A69000; PMID:98037514; PMID:93711463
 A;Accession: H69194
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-311 <MTH>
 A;Cross-references: GB:AE000850; GB:AE000666; NID:G2621794; PIDN:AAE85215.1; PID:G262179
 A;Experimental source: strain Delta H
 C;Genetics:
 A;Gene: MTH710
 A;Start codon: GTG

Query Match 62.5%; Score 35; DB 2; Length 311;
 Best Local Similarity 63.6%; Pred. No. 27;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 EEVVPXGMHYS 11
 |||||

Db 219 EEVVGSLHES 229

RESULT 10
 T24111
 hypothetical protein R10D12.10 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C;Accession: T24111
 R;Percy, C.
 Submitted to the EMBL Data Library, October 1996
 A;Reference number: Z19842
 A;Accession: T24111
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-425 <WIL>
 A;Cross-references: EMBL:Z81109; PIDN:CAB03241.1; GSPDB:GN00023; CESP:R10D12.10
 A;Experimental source: clone R10D12
 C;Genetics:
 A;Gene: CESP:R10D12.10
 A;Map position: 5
 A;Intons: 23/3; 56/3; 113/3; 257/2

Query Match 62.5%; Score 35; DB 2; Length 425;
 Best Local Similarity 50.0%; Pred. No. 37;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 EEVVPXGMH 10
 |||||
 Db 335 EQVPGGLQY 344

RESULT 11
 G86430
 T518.1 protein - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002
 C;Accession: G86430
 R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alons
 Chn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K
 ansen, N.F.; Hughes, B.; Hultzar, L.
 Nature 408, 816-820, 2000
 A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzai
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A;Reference number: A86141; PMID:21016719; PMID:11130712
 A;Accession: G86430
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-510 <SPO>
 A;Cross-references: GB:AE005172; NID:G4587512; PIDN:AAD25743.1; GSPDB:GN00141
 C;Genetics:
 A;Map position: 1
 C;Superfamily: hexose phosphate transport protein uhpT

Query Match 62.5%; Score 35; DB 2; Length 510;
 Best Local Similarity 60.0%; Pred. No. 45;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 EEVVPXGMH 10
 |||||
 Db 12 EEVVPGLIHF 21

RESULT 12
 G69117
 diaphytine synthase - Methanobacterium thermoautotrophicum (strain Delta H)
 C;Species: Methanobacterium thermoautotrophicum
 C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 17-Feb-2003
 C;Accession: G69117

R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Qiu, D.; Spadafora, R.; Vitcure, R.; Mang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
ki, S.; Church, G.M.; Daniels, C.D.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: func
A;Reference number: A69000; MUID:98037514; PMID:93711463
A;Accession: G69117
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-264 <MTH>
A;Cross-references: GB:AB000940; GB:AE000666; NID:G2623011; PIDN:AMB6340.1; PID:G262301
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH1874

Query Match 60.7%; Score 34; DB 2; Length 264;
Best Local Similarity 62.5%; Pred. No. 35;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMHY 10
DB 235 VVPAGLHF 242

RESULT 13
C75538
hypothetical protein - *Deinococcus radiodurans* (strain R1)
C;Species: *Deinococcus radiodurans*
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
C;Accession: C75538
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: C75538
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-279 <WHI>
A;Cross-references: GB:AE001889; GB:AE000513; NID:G6457944; PIDN:AAF09867.1; PID:G645795
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0271
A;Map position: 1
C;Superfamily: *Deinococcus radiodurans* hypothetical protein DR0271

Query Match 60.7%; Score 34; DB 2; Length 279;
Best Local Similarity 75.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 VPXGMHY 11
DB 100 VPLGRHY 107

RESULT 14
B75478
3-dehydroquinate synthase - *Deinococcus radiodurans* (strain R1)
C;Species: *Deinococcus radiodurans*
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: B75478
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: B75478
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-350 <WHI>
A;Cross-references: GB:AE001932; GB:AB000513; NID:G6458481; PIDN:AAF10353.1; PID:G645848

A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0777
A;Map position: 1
C;Superfamily: 3-dehydroquinate synthase; 3-dehydroquinate synthase homology

Query Match 60.7%; Score 34; DB 2; Length 350;
Best Local Similarity 60.0%; Pred. No. 48;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGMHY 11
DB 252 EAVVAGMHYA 261

RESULT 15
T35025
probable DNA ligase - *Streptomyces coelicolor*
C;Species: *Streptomyces coelicolor*
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C;Accession: T35025
R;Seeger, S.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, June 1999
A;Reference number: Z21565
A;Accession: T35025
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-355 <SEB>
A;Cross-references: EMBL:AL079355; PIDN:CAB45581.1; GSPDB:GN00070; SCODEB:SC4C6.17C
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCODEB:SC4C6.17C

Query Match 60.7%; Score 34; DB 2; Length 355;
Best Local Similarity 71.4%; Pred. No. 49;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 VVPXGMHY 10
DB 20 IPXGMHY 26

Search completed: December 22, 2003, 17:44:55
Job time : 12.0667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:43:51 ; Search time 4.6 Seconds

(without alignment)
112.455 Million cell updates/sec

Title: US-09-909-164-7
Perfect score: 56
Sequence: 1 EEVVPXGMHYS 11

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	67.9	1058	1	CARB_FUSNN
2	37	66.1	1057	1	AROQ_VIBPA
3	37	66.1	1057	1	CARB_STAM
4	37	66.1	1057	1	CARB_STAM
5	36	64.3	102	1	YIAK_TYDA
6	36	64.3	460	1	YMEI_YEAST
7	36	64.3	743	1	YK47_YEAST
8	35	62.5	227	1	ID11_MESAU
9	35	62.5	308	1	GAAB_METH
10	34	60.7	426	1	GAAB_METH
11	33	58.9	441	1	YI15_VIBPA
12	33	58.9	513	1	PHSL_DESBA
13	33	58.9	627	1	MUTL_BACSU
14	33	58.9	1188	1	KPBA_CABEL
15	33	58.9	1396	1	VITF_BPTS
16	32	57.1	126	1	TKMK_BOVIN
17	32	57.1	233	1	RS2_CLOAB
18	32	57.1	267	1	RR2_CHLV
19	32	57.1	288	1	CGD2_RAT
20	32	57.1	289	1	CGD2_HUMAN
21	32	57.1	289	1	CGD2_MOUSE
22	32	57.1	291	1	CGD1_BRARE
23	32	57.1	291	1	CGD1_XENLA
24	32	57.1	291	1	CGD2_CHICK
25	32	57.1	291	1	CGD2_CHICK
26	32	57.1	292	1	CGD1_CHICK
27	32	57.1	292	1	CGD3_HUMAN
28	32	57.1	295	1	CGD1_HUMAN
29	32	57.1	295	1	CGD1_MOUSE
30	32	57.1	295	1	CGD1_MOUSE
31	32	57.1	341	1	HYPE_AZOVI
32	32	57.1	353	1	T3BA_BACAR
33	32	57.1	573	1	SUOX_DROME

34	32	57.1	578	1	MDLB_BUCCP	Q89A96 buchnera ap
35	32	57.1	759	1	SC71_YEAST	P37784 saccharomyc
36	32	57.1	877	1	SULH_SCHRO	O74377 echizosacch
37	32	57.1	1401	1	REOC_VIBCH	Q94V29 vibrio chol
38	32	57.1	2717	1	KEP1_HUMAN	P15822 homo sapien
39	31.5	56.2	847	1	CD22_HUMAN	P20273 homo sapien
40	31	55.4	124	1	REV_SIVCZ	P17280 chimpanzee
41	31	55.4	130	1	SZ05_RAT	P97885 rattus norv
42	31	55.4	276	1	V939_METJA	O58349 methanococc
43	31	55.4	319	1	YHAI_CRYPA	P10941 cryphonectr
44	31	55.4	331	1	RJ3_ARCFU	O28354 archaeoglob
45	31	55.4	363	1	ALFB_SHSEP	P52210 ovib aries

ALIGNMENTS

RESULT 1	CARB_FUSNN	STANDARD	PRT: 1058 AA.
AC	O8RG86		
DT	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (carbamoyl-phosphate synthetase ammonia chain).		
GN	CARB OR F00422.		
OS	Fusobacterium nucleatum (subsp. nucleatum).		
OC	Bacteria; Fusobacteriales; Fusobacteriaceae; Fusobacterium.		
OX	NCBI_TaxID=76856;		
RY	[1]		
RP	SEQUENCE FROM N. A.		
RC	STRAIN=ATCC 25586;		
FX	MEDLINE=21886394; PubMed=11889109;		
RA	Kapactral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A., Bhattacharyya A., Barman A., Gardner W., Grechkin G., Zhu L., Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A., Larsen N., D'Souza M., Malinas T., Fusch G., Haselkorn R., Forststein M., Kyridis N., Overbeek R.;		
RT	"Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586."		
J	Bacteriol. 184:2005-2018(2002).		
-1	CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.		
-1	COFACTOR: Binds 3 manganese ions per subunit (By similarity).		
-1	PATHWAY: Pyrimidine biosynthesis.		
-1	SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).		
-1	SIMILARITY: BELONGS TO THE CARB FAMILY.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_email_to_license@isb-sib.ch).		
CC	EMBL; AEO10554; ALA94625.1; ALT_INT.		
DR	HAMAP; MF_01210; -1.		
DR	InterPro; IPR006275; CARA_L_glu.		
DR	InterPro; IPR005483; CPase_L.		
DR	InterPro; IPR005479; CPase_L_D2.		
DR	InterPro; IPR005480; CPase_L_D3.		
DR	InterPro; IPR005481; CPase_L_N.		
DR	InterPro; IPR004362; MGS_1ike.		
DR	Pfam; PF00289; CPSase_L_chain; 2.		
DR	Pfam; PF02786; CPSase_L_D2; 2.		
DR	Pfam; PF02787; CPSase_L_D3; 1.		

```

DR PFam: PF02142; MGS: 1.
DR PRINTS; PR00098; CPSASE.
DR TIGRFAMs; TIGR01369; CPSase1_1; 1.
DR PROSITE; PS00866; CPSASE_1; 2.
DR PROSITE; PS00867; CPSASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
  AMP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1058
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF7C1E39F CRC64;

Query Match 67.9%; Score 38; DB 1; Length 1058;
Best Local Similarity 60.0%; Pred. No. 10;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EYVFXGMHYS 11
Db 190 EIVVGNLNTS 199

RESULT 2
AROQ_VIBPA STANDARD; PRT; 426 AA.
ID AROQ_VIBPA
AC 0870X9;
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
  enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPs).
GN AROQ OR VP1020.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
  Vibrionaceae; Vibrrio.
OX NCBI_TaxID=670;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=RTMD 221063 / Serotype O3:K6;
RX MEDLINE=22508454; Pubmed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
  Iijima Y., Najiima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
  Yaenaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
  "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
  distinct from that of V. cholerae.";
RT Lancet 361:743-749(2003).
RL
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
  phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
  sixth step.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the EPSP synthase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
CC EMBL; AP005076; BACS9283.1; -
DR HAMAP; MF_00210; -; 1.

```

```

DR PROSITE; PS00104; EPSP SYNTHASE 1; 1.
DR PROSITE; PS00885; EPSP SYNTHASE 2; 1.
KW Aromatic amino acid biosynthesis; Transferase; Complete proteome.
SQ SEQUENCE 426 AA; 46094 MW; 373D39CC5BA1F70F CRC64;

Query Match 66.1%; Score 37; DB 1; Length 426;
Best Local Similarity 60.0%; Pred. No. 6,3;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EYVFXGMHYS 10
Db 223 EYVFXGMHYS 232

RESULT 3
CARB_STAM STANDARD; PRT; 1057 AA.
ID CARB_STAM
AC 099UR5;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase ammonia chain.
DE Carbamoyl-phosphate synthase ammonia chain.
GN CARB OR PYRAB OR SAV1203 OR SAV1046.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
  Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=Mu50 / ATCC 700699, and N315;
RX MEDLINE=21311952; Pubmed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
  Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-Q., Ito T.,
  Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
  Mizunuma U. Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
  Sekimizu K., Hirakawa H., Kohara S., Goto S., Yabuzaki J.,
  Kanetsuna M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
  Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
  "Whole genome sequencing of methicillin-resistant Staphylococcus
  aureus.";
RT Lancet 357:1225-1240(2001).
RL
CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
  phosphate + L-glutamate + carbamoyl phosphate.
CC -1- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -1- PATHWAY: Arginine biosynthesis.
CC -1- PATHWAY: Pyrimidine biosynthesis; first step.
CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain
  promotes the hydrolysis of glutamine to ammonia, which is used by
  the large (or ammonia) chain to synthesize carbamoyl phosphate (By
  similarity).
CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
CC EMBL; AP003361; BAB57365.1; -
DR EMBL; AP003132; BAB42298.1; -
DR PIR; F89892; F89892.
DR HSP; P00968; IC50.
DR HAMAP; MF_01210; -; 1.
DR InterPro; IPR006275; CARB_L_glu.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR004362; MGS like.
DR PFam; PF00289; CPSase_L_chain; 2.

```

```

DR pfam; PF02786; CPSase_L_D2; 2.
DR pfam; PF02787; CPSase_L_D3; 1.
DR pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPSASE.
DR TIGRPFAMs; TIGR01369; CPSase1_1rg; 1.
DR PROSITE; PS00866; CPSASE_1; 2.
DR PROSITE; PS00867; CPSASE_2; 2.
DR Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 529 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1057 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1057
FT NP_BIND 153 210
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 ATP (POTENTIAL).
FT METAL 298 298 MANGANESE 1 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1057 AA; 117171 MW; E3E179EF051F0F8 CRC64;

Query Match 66.1%; Score 37; DB 1; Length 1057;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 EVVFXGMHYS 11
Db 190 EIVSNGHYS 199

RESULT 4
CARB STAM STANDARD; PRT; 1057 AA.
ID CARB STAM STANDARD; PRT; 1057 AA.
AC P56940;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
phosphate synthetase ammonia chain).
GN CARB OR PYRAB OR MM1086.
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=196620;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naito T., Kuroda H., Cui L.,
RA Yamamoto K., Hiratsuka K.;
RT "Genome and virulence determinants of high virulence community-
acquired MRSA."
RL Lancet 359:1819-1827(2002).
CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
phosphate + L-glutamate + carbamoyl phosphate.
CC -1- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -1- PATHWAY: Pyrimidine biosynthesis; first step.
CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain
promotes the hydrolysis of glutamine to ammonia, which is used by
the large (or ammonia) chain to synthesize carbamoyl phosphate (By
similarity).
CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.ebi.ac.uk/announcements
or send an email to license@ebi.ac.uk).
```

```

CC EMBL; AP004825; BAB94951.1; -.
DR HAMAP; MF_01210; -.
DR InterPro; IPR006275; CARB_L_glu.
DR InterPro; IPR005483; CPSase_L.
DR InterPro; IPR005479; CPSase_L_D2.
DR InterPro; IPR005480; CPSase_L_D3.
DR InterPro; IPR005481; CPSase_L_N.
DR InterPro; IPR004362; MGS_1like.
DR pfam; PF00289; CPSase_L_Chain; 2.
DR pfam; PF02786; CPSase_L_D2; 2.
DR pfam; PF02787; CPSase_L_D3; 1.
DR pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPSASE.
DR TIGRPFAMs; TIGR01369; CPSase1_1rg; 1.
DR PROSITE; PS00866; CPSASE_1; 2.
DR PROSITE; PS00867; CPSASE_2; 2.
DR Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1057 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1057
FT NP_BIND 153 210
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 ATP (POTENTIAL).
FT METAL 298 298 MANGANESE 1 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1057 AA; 117185 MW; DBE3B09F86CF152 CRC64;

Query Match 66.1%; Score 37; DB 1; Length 1057;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 EVVFXGMHYS 11
Db 190 EIVSNGHYS 199

RESULT 5
YLIK TYDVA STANDARD; PRT; 102 AA.
ID YLIK TYDVA STANDARD; PRT; 102 AA.
AC P31619;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Hypothetical 11.2 kDa protein.
GN V1.
OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
OC Viruses; ssDNA viruses; Geminiviridae; Mastreviruses.
OX NCBI_TaxID=31599;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92188538; PubMed=1546458;
RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.B.;
RT "The nucleotide sequence of the infectious cloned DNA component of
tobacco yellow dwarf virus reveals features of geminiviruses
infecting monocotyledonous plants."
RL Virology 187:633-642(1992).
CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.ebi.ac.uk/announcements
or send an email to license@ebi.ac.uk).
```

```

DR PIR: A42452; A42452.
DR InterPro; IPR002621; Gemini_mov.
DR Pfam; PF01708; Gemini_mov; 1.
DR Hypothetical protein.
SQ SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF55B67 CRC64;

Query Match      64.3%; Score 36; DB 1; Length 102;
Best Local Similarity 60.0%; Pred. No. 2.3;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Oy 2 EVVPGMHY 11
Db 7 QVPSGIVYS 16

RESULT 6
UMEL_YEAST STANDARD; PRT; 460 AA.
ID UMEL_YEAST
AC Q03010; P87330; (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Meiosis negative regulator UME1.
GN UME1 OR WTM3 OR YPL139C OR LPI7C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=AS364A;
RA Mallory M.J., Strich R.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=SS288C / AB972;
RX MEDLINE=97313271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,
RA Araujo R., Aparicio A., Barrett B.G., Badcock K., Benes V.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Dueseteroelt A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,
RA Hunnicke-Smith S., Hyman R., Johnston M., Kaiman S., Kleene K.,
RA Komp C., Kurd O., Lashkar D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mitalpali S., Pearson D.,
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA Petel F.X., Pohl T.M., Fumelle D., Schater M., Scharte M.,
RA Scherens B., Schramm S., Schroeder W., Sdicu A.M., Tettelin H.,
RA Uristarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA Walsh S.V., Wambut R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Han J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI."
RL Nature 387:103-105(1997).
CC -1- FUNCTION: TRANSCRIPTIONAL MODULATOR WITH ROLES IN MEIOTIC
CC REGULATION AND SILENCING. NEGATIVE REGULATOR OF MEIOSIS.
CC -1- SIMILARITY: Contains 4 WD repeats.
CC -1- SIMILARITY: STRONG, TO YEAST WTM1 AND WTM2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U10280; AAB40937.1; -
CC EMBL; U43703; AAB68221.1; -
DR PIR; S69046; S69046.
DR TRANSFAC; T04309; -
DR SGD; S0006060; UME1.

```

```

DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0003714; F:transcription co-repressor activity; IDA.
DR GO; GO:0040020; P:regulation of meiosis; IGI.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 3.
DR SMART; SM00320; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS_1; FALSE_NEG.
DR PROSITE; PS00682; WD_REPEATS_2; FALSE_NEG.
DR PROSITE; PS50294; WD_REPEATS_REGION; FALSE_NEG.
KW Transcription regulation; Meiosis; Repeat; WD repeat.
FT REPEAT 233 271 WD 1.
FT REPEAT 276 316 WD 2.
FT REPEAT 339 379 WD 3.
FT REPEAT 411 451 WD 4.
SQ SEQUENCE 460 AA; 51022 MW; AA6F604487BCBA9 CRC64;

Query Match      64.3%; Score 36; DB 1; Length 460;
Best Local Similarity 62.5%; Pred. No. 11;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 3 VVPXGMHY 10
Db 85 IVPGLAHY 92

RESULT 7
YK47_YEAST STANDARD; PRT; 743 AA.
ID YK47_YEAST
AC P36148;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 83.6 kDa protein in CCP1-MET1 intergenic region.
GN YKR067W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=SS288C;
RA van Vlier-Reedijk J.C., Planta R.J.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: STRONG, TO YEAST YBL011W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z28292; CAA82146.1; -
DR PIR; S38143; S38143.
DR SGD; S0001775; GPT2.
DR GO; GO:0005753; C:endoplasmic reticulum; IDA.
DR GO; GO:0004366; F:glucocorticoid-3-phosphate O-acetyltransferase acti. .; IDA.
DR GO; GO:0008654; P:phospholipid biosynthesis; IDA.
DR InterPro; IPR00123; Acyltransferase.
DR Pfam; PF01553; Acyltransferase; 1.
DR SMART; SM00563; P18C; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 31 55 POTENTIAL.
FT TRANSMEM 69 85 POTENTIAL.
FT TRANSMEM 502 524 POTENTIAL.
FT TRANSMEM 539 555 POTENTIAL.
SQ SEQUENCE 743 AA; 83644 MW; 84B9946B56B82F15 CRC64;

Query Match      64.3%; Score 36; DB 1; Length 743;
Best Local Similarity 75.0%; Pred. No. 18;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

Qy 3 VEPXGMHY 10
 Db 294 VEPXGMHY 301

RESULT 8

ID11 MESAU STANDARD; PRT: 227 AA.

AC 035566; 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Isopentenyl-diphosphate delta-isomerase 1 (EC 5.3.3.2) (IPP isomerase 1) (isopentenyl pyrophosphate isomerase 1) (IPPII).
 GN ID11.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OC NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9373600; PubMed=9228075;
 RA Paton V.G., Shackelford J.E., Kirsans S.K.;
 RT "Cloning and subcellular localization of hamster and rat isopentenyl
 RT diphosphate dimethylallyl diphosphate isomerase. A PstI motif targets
 RT the enzyme to peroxisomes.";
 RL J. Biol. Chem. 272:18945-18950(1997).
 CC -1- FUNCTION: CATALYZES THE 1,3-ALLYLIC REARRANGEMENT OF THE
 CC HOMOLALYLIC SUBSTRATE ISOPENTENYL (IPP) TO ITS HIGHLY
 CC ELECTROPHILIC ALLYLIC ISOMER, DIMETHYLLALYL DIPHOSPHATE (DMAPP).
 CC -1- CATALYTIC ACTIVITY: Isopentenyl diphosphate = dimethylallyl
 CC diphosphate.
 CC -1- COFACTOR: REQUIRES MAGNESIUM FOR ACTIVITY.
 CC -1- PATHWAY: ISOPRENOID BIOSYNTHETIC PATHWAY WHOSE END PRODUCTS
 CC INCLUDE DOLOCHOLS, VITAMINS A, D, E, AND K, STEROID HORMONES,
 CC CAROTENOIDS BILE ACIDS AND CHOLESTEROL.
 CC -1- SUBCELLULAR LOCATION: Peroxisomal.
 CC -1- SIMILARITY: Belongs to the IPP isomerase type 1 family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC
 CC EMBL: AF003836; AAC53283.1; -
 DR InterPro: IPR002667; IPP_isomerase.
 DR InterPro: IPR000086; NUDIX_hydrolase.
 DR Pfam: PF00293; NUDIX; 1.
 DR ProDom: PD004109; IPP_isomerase; 1.
 KM Carotenoid biosynthesis; Cholesterol biosynthesis;
 KM Isoprene biosynthesis; Sterol biosynthesis; Isomerase; Peroxisome;
 KM Magnesium.
 FT ACT SITE 86
 FT ACT SITE 86
 FT ACT SITE 148
 FT SITE 225
 FT SITE 227
 SQ SEQUENCE 227 AA; 26317 MM; F500A6586385E803 CRC64;

Query Match 62.5%; Score 35; DB 1; Length 227;
 Best Local Similarity 70.0%; Pred. No. 8.4;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EEVVPXGMHY 10
 Db 121 EEVDPNEMHY 130

RESULT 9
 GAAB_METH

ID GAAB METH STANDARD; PRT: 308 AA.

AC 026806; 15-OCT-2001 (Rel. 40, Created)
 DT 15-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE GMP synthase [glutamine-hydrolyzing] subunit B (EC 6.3.5.2) (GMP
 DE synthetase).
 GN GAAB OR MTH710.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 OC NCBI_TaxID=187420;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Delta H;
 RX MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
 RA Alredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pochler B., Qiu D.,
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shiner G., Goyal A., Pietrovski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT deltaH: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155(1997).
 CC -1- CATALYTIC ACTIVITY: ATP + xanthosine 5'-phosphate + L-glutamine +
 CC H(2)O = AMP + diphosphate + GMP + L-glutamate.
 CC -1- PATHWAY: GMP biosynthesis.
 CC -1- SUBUNIT: HETERODIMER COMPOSED OF A GLUTAMINE AMIDOTRANSFERASE
 CC SUBUNIT (A) AND A GMP SYNTHASE SUBUNIT (B) (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE GMP SYNTHASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC
 CC EMBL: AE000850; AAB85215.1; ALT_INIT.
 DR HSSP: P04079; IGPM.
 DR HAMAP: MF_00345; -; 1
 DR InterPro: IPR001674; GMP_synth_C.
 DR Pfam: PF00958; GMP_synth_C; 1.
 DR TIGRFAMs: TIGR00884; guaA_Cterm; 1.
 KM Ligase; GMP biosynthesis; Purine biosynthesis; ATP-binding;
 KM Complete proteome.
 FT DOMAIN 33 184
 FT NP BIND 29 35
 FT NP BIND 35
 SQ SEQUENCE 308 AA; 34403 MM; F2DCF6ED202CABC1 CRC64;

Query Match 62.5%; Score 35; DB 1; Length 308;
 Best Local Similarity 63.6%; Pred. No. 12;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EEVVPXGMHY 11
 Db 216 EEVVEGLHES 226

RESULT 10
 SLST_YARLI STANDARD; PRT: 426 AA.

AC 099158; 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE SLST protein precursor.
 GN SLST.
 OS Yarrowia lipolytica (Candida lipolytica).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Dipodacaceae; Yarrowia.
 RX NCBI_TaxID=4952;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 20460 / W29;
 RA MEDLINE=96216076; PubMed=662639;
 RA Bolesame A., Beckerich J.-M., Galliard C.;
 RT "Is1p, an endoplasmic reticulum component, is involved in the
 RT protein translocation process in the yeast Yarrowia lipolytica";
 RL J. Biol. Chem. 271:11668-11675 (1996).
 CC -1- FUNCTION: INVOLVED IN THE PROTEIN TRANSLATION PROCESS. MAY
 CC INTERACT DIRECTLY WITH TRANSLATING POLYPEPTIDES TO FACILITATE
 CC THEIR TRANSFER AND/OR HELP THEIR FOLDING IN THE ER. IT IS NOT
 CC REQUIRED FOR VIABILITY BUT IS ESSENTIAL FOR OPTIMAL GROWTH AT
 CC ELEVATED TEMPERATURES.
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z50154; CA90516.1; -.
 DR PIR; S58132; S58132.
 DR InterPro; IPR000866; ER_target.
 DR PROSITE; PS00014; ER_TARGET; 1.
 KW Endoplasmic reticulum; Signal.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 426 SL1 PROTEIN.
 FT SITE 423 426 PREVENT SECRETION FROM ER (POTENTIAL).
 SQ SEQUENCE 426 AA; 47201 MW; 0ACDDEP17540B8E2 CRC64;
 Query Match 60.7%; Score 34; DB 1; Length 426;
 Best Local Similarity 44.4%; Pred. No. 26;
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EEVVPXGMX 9
 DB 52 DOVTPAGLH 60
 RESULT 11
 ID YL15_VIBPA STANDARD; PRT; 441 AA.
 AC P46231;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hypothetical protein VP2115 (ORF3).
 GN VP2115.
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrrio.
 OX NCBI_TaxID=670;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIMD 221063 / Serotype O3:K6;
 RA MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Nishina M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yaenaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
 RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
 RT distinct from that of V. cholerae.";
 RL Lancet 361:743-749 (2003).
 CC [2]
 CC SEQUENCE OF 1-140 FROM N.A.
 RC STRAIN=BB22;
 RA MEDLINE=94292449; PubMed=8021208;
 RA McCarter L.L.;
 RT "MotY, a component of the sodium-type flagellar motor.";

RL J. Bacteriol. 176:4219-4225 (1994).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: STRONG, TO H. INFLUENZAE HI0325.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AP005080; BAC60378.1; -.
 DR EMBL; U06949; AAA21571.1; -.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 21 41 POTENTIAL.
 FT TRANSMEM 51 71 POTENTIAL.
 FT TRANSMEM 94 114 POTENTIAL.
 FT TRANSMEM 118 138 POTENTIAL.
 FT TRANSMEM 150 170 POTENTIAL.
 FT TRANSMEM 195 215 POTENTIAL.
 FT TRANSMEM 239 259 POTENTIAL.
 FT TRANSMEM 260 280 POTENTIAL.
 FT TRANSMEM 291 311 POTENTIAL.
 FT TRANSMEM 334 354 POTENTIAL.
 FT TRANSMEM 363 383 POTENTIAL.
 FT TRANSMEM 419 439 POTENTIAL.
 SQ SEQUENCE 441 AA; 45961 MW; 451969FE307B4D46 CRC64;
 Query Match 58.9%; Score 33; DB 1; Length 441;
 Best Local Similarity 54.5%; Pred. No. 43;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 EEVVPXGMXYS 11
 DB 417 ETVPPTFIHYN 427
 RESULT 12
 ID PHSU_DSBSA STANDARD; PRT; 513 AA.
 AC P13065;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Periplasmic (NiFeSe) hydrogenase large subunit (EC 1.12.99.6) (NiFeSe
 DE hydrogenylase large chain).
 OS Desulfovibrio baculatus (Desulfomicrobium baculatus).
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
 OC Desulfomicrobiaceae; Desulfomicrobium.
 OX NCBI_TaxID=899;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=88058744; PubMed=3316183;
 RA Menon N.K., Peck H.D. Jr., Je Gall J., Przybyla A.E.;
 RT "Cloning and sequencing of the genes encoding the large and small
 RT subunits of the periplasmic (NiFeSe) hydrogenase of Desulfovibrio
 RT baculatus.";
 RL J. Bacteriol. 169:5401-5407 (1987).
 CC [2]
 CC REVISIONS.
 RA Menon N.K., Peck H.D. Jr., Je Gall J., Przybyla A.E.;
 RL J. Bacteriol. 170:4429-4429 (1988).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS).
 RA MEDLINE=9306038; PubMed=10378275;
 RA Garcin E., Vernede X., Hatchikian E.C., Volbeda A., Frey M.,
 RA Pontecilla-Camps J.C.;
 RT "Removal of the bridging ligand atom at the Ni-Fe active site of
 RT [NiFe] hydrogenase upon reduction with H2, as revealed by X-ray
 RT structure analysis at 1.4-A resolution.";
 RL Structure 7:557-566 (1999).
 CC -1- CATALYTIC ACTIVITY: H(2) + acceptor = 2 H(+) + reduced acceptor.

CC -1- COFACTOR: Nickel, 2 Irons and selenocysteine. Iron 1 has three
CC cyanide and carbon monoxide ligands. Iron 2 has three water
CC ligands.
CC -1- SUBUNIT: Heterodimer of a large and a small subunit.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- MISCELLANEOUS: PERHAPS THE LEADER OF THE SMALL SUBUNIT SERVES AS A
CC TRANSPORT VEHICLE FOR BOTH SUBUNITS.
CC -1- SIMILARITY: BELONGS TO THE [NIFE]/[NIFSE] HYDROGENASE LARGE
CC SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; M18271; AAA23375.2; -.
CC PIR; A33101; HDVBLB.
CC PDB; 1CC1; 01-JUN-99.
CC InterPro: IPR001501; N1_hdl.
CC Pfam; PF00374; NIFSE_Haees; 1.
CC PROSITE; PS00507; N1_HGENSE_L_1; 1.
CC PROSITE; PS00508; N1_HGENSE_L_2; 1.
CC Oxidoreductase; Periplasmic; Metal-binding; Nickel; Iron; Selenium;
KM Selenocysteine; 3D-structure.
FT INIT MET 0 0
FT METAL 51 51 IRON 2.
FT METAL 70 70 NICKEL.
FT METAL 73 73 IRON 1.
FT METAL 73 73 NICKEL.
FT METAL 444 444 IRON 2 (VIA CARBONYL OXYGEN).
FT METAL 492 492 NICKEL.
FT METAL 495 495 IRON 1.
FT METAL 495 495 NICKEL.
FT METAL 498 498 IRON 2.
FT SE_CYS 492 492
FT STRAND 13 16
FT STRAND 22 23
FT STRAND 26 33
FT TURN 34 35
FT STRAND 36 44
FT STRAND 46 46
FT STRAND 50 53
FT HELIX 54 56
FT TURN 54 56
FT HELIX 59 61
FT HELIX 62 65
FT HELIX 66 69
FT HELIX 74 89
FT TURN 90 90
FT HELIX 95 119
FT TURN 120 121
FT HELIX 122 124
FT TURN 125 125
FT TURN 132 133
FT TURN 139 140
FT HELIX 144 147
FT HELIX 151 183
FT STRAND 192 193
FT STRAND 194 195
FT STRAND 196 197
FT HELIX 203 222
FT TURN 223 223
FT HELIX 224 234
FT HELIX 236 239
FT STRAND 248 250
FT STRAND 253 255
FT TURN 258 259
FT STRAND 263 265
FT STRAND 268 271
FT TURN 272 273
FT STRAND 274 276

FT HELIX 280 282
FT STRAND 283 286
FT TURN 288 289
FT STRAND 290 290
FT TURN 291 291
FT STRAND 292 292
FT STRAND 300 301
FT HELIX 302 304
FT STRAND 309 309
FT TURN 311 312
FT TURN 314 315
FT STRAND 318 318
FT STRAND 322 325
FT TURN 326 327
FT STRAND 328 328
FT STRAND 331 331
FT HELIX 334 341
FT STRAND 344 344
FT STRAND 347 357
FT STRAND 363 363
FT HELIX 364 371
FT HELIX 374 398
FT TURN 401 402
FT STRAND 406 406
FT STRAND 415 423
FT TURN 424 425
FT STRAND 426 435
FT TURN 436 437
FT STRAND 438 446
FT TURN 447 447
FT HELIX 448 451
FT TURN 457 458
FT HELIX 463 468
FT TURN 469 470
FT STRAND 472 472
FT TURN 476 477
FT HELIX 480 488
FT TURN 489 489
FT STRAND 492 492
FT HELIX 493 497
SQ SEQUENCE 513 AA; 56683 MW; AC8285A6F80576FC CRC64;
Query Match 58.9%; Score 33; DB 1; Length 513;
Best Local Similarity 71.4%; Pred. No. 50;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 5 PKGMYS 11
DB 297 PGGLHYS 303
RESULT 13
MOTL_BACSU STANDARD; PRT; 627 AA.
ID MOTL_BACSU
AC P49850;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA mismatch repair protein mutL.
GN MOTL.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96349107; PubMed=8760914;
RA Ginetti F., Perego M., Albertini A.M., Galizzi A.;
RT "Bacillus subtilis mutL muts mult operon: identification, nucleotide
sequence and mutagenesis."; Microbiology 142:2021-2029(1996).
RL [2]
RP SEQUENCE FROM N.A.

RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borries R., Boursier L., Brans A., Braun M., Briggell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabre C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holstappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kaashwa Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Meliadi R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Pressecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Solido B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Torachi M., Tanakoshi A., Tanaka T., Terpatra P., Tognoni A.,
 RA Tostato V., Uchiyama S., Vandenbol M., Vannier J., Vassarotti A.,
 RA Viari A., Wandut R., Wedler E., Wedler H., Weitzenger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumestein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT subtilis";
 RL Nature 390:249-256(1997).
 CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES IN
 CC DNA. IT IS REQUIRED FOR DAM-DEPENDENT METHYL-DIRECTED DNA MISMATCH
 CC REPAIR. MAY ACT AS A "MOLECULAR MATCHMAKER". A PROTEIN THAT
 CC PROMOTES THE FORMATION OF A STABLE COMPLEX BETWEEN TWO OR MORE
 CC DNA-BINDING PROTEINS IN AN ATP-DEPENDENT MANNER WITHOUT ITSELF
 CC BEING PART OF A FINAL EFFECTOR COMPLEX (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MOTIF/HEX FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U27343; AAA19236.1; -;
 DR EMBL: Z99112; CAB13578.1; -;
 DR PIR: A69663; A69663.
 DR HSP: P23567; IBKN.
 DR Subtilist; BG11402; mutL.
 DR HAMAP; MF_00149; -; 1.
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR002099; DNA_mis_repair.
 DR Pfam; PF01119; DNA_mis_repair; 1.
 DR Pfam; PF02518; HATPase_C; 1.
 DR SMART; SM00387; HATPase_C; 1.
 DR TIGRFAMs; TIGR00585; mutL; 1.
 DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
 KW DNA repair; Complete proteome.
 SQ SEQUENCE 627 AA; 70431 MW; 068A0509CC265343 CRC64;

Query Match 58.9%; Score 33; DB 1; Length 627;
 Best Local Similarity 54.5%; Pred. No. 62;
 Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 BEVVPXGMHYS 11
 DB 488 EMIVFLTRHYS 498

RESULT 14
 ID_KPBA_CAEEL STANDARD; PRT; 1188 AA.
 AC P34335;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable phosphotyrase B kinase alpha regulatory chain (Phosphotyrase
 DE kinase alpha subunit).
 GN C14B9.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Briceol N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kizeten J., Laiseter N.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterston R., Watson A., Wellstock L., Wilkinston-Sproat J.,
 RA Wholden P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.*
 RT elegans";
 RL Nature 368:32-38(1994).
 RN [2]
 RP REVISIONS.
 RA Waterston R.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PHOSPHORYLASE B KINASE CATALYZES THE PHOSPHORYLATION OF
 CC SERINE IN CERTAIN SUBSTRATES, INCLUDING TROPONIN I. THE ALPHA
 CC CHAIN MAY BIND CALMODULIN (BY SIMILARITY).
 CC -1- PATHWAY: Glycogen metabolism.
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHORYLASE B KINASE REGULATORY
 CC CHAINS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L15188; AAA27954.2; -;
 DR WormPep; C14B9.8; CE26870.
 KW Hypothetical protein; Glycogen metabolism; Calmodulin-binding.
 SQ SEQUENCE 1188 AA; 135558 MW; D83BB875F3603863 CRC64;

Query Match 58.9%; Score 33; DB 1; Length 1188;
 Best Local Similarity 54.5%; Pred. No. 1.2e+02;
 Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 BEVVPXGMHYS 11
 DB 950 BEVMDGIRHYS 960

RESULT 15
 ID_VLTF_BPT5 STANDARD; PRT; 1396 AA.
 AC P13350; O48502;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)

```

DE L-shaped tail fiber protein (LTF protein).
GN LTF.
OS Bacteriophage T5.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC T5-like viruses.
OX NCBI_TaxID=10726;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95309401; PubMed=7789514;
RA Kaliman A.V., Kulehin V.E., Shlyapnikov M.G., Ksenzenko V.N.,
RT "The nucleotide sequence of the bacteriophage T5 ltf gene.";
RL FEBS Lett. 366:46-48(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Kaliman A.V.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=88289370; PubMed=3267228;
RA Kaliman A.V., Kryukov V.M., Bayev A.A.;
RT "The nucleotide sequence of bacteriophage T5 DNA at the region
RL between early and late genes.";
RL Nucleic Acids Res. 16:6230-6230(1988).
CC -1- FUNCTION: NONESSENTIAL PROTEIN THAT MEDIATES BINDING TO THE
CC POLYMANNOSE O ANTIGEN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X69460; CAA49220.1; -.
DR EMBL; AJ001191; CAA04591.1; -.
DR PIR; S65934; S36851.
KM Late protein.
FT CONFLICT 986 V -> A (IN REF. 2).
SQ SEQUENCE 1396 AA; 147989 MW; 18CD2192F65FFFC1 CRC64;

```

Query Match 58.9%; Score 33; DB 1; Length 1396;
 Best Local Similarity 50.0%; Pred. No. 1.4e+02;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

QY 2 EVVFXGMHYS 11
   :|||
Db 1360 KTIPADPHYS 1369

```

Search completed: December 22, 2003, 17:42:24
 Job time : 5.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:27:26 ; Search time 25.2 Seconds

(without alignments)
112.642 Million cell updates/sec

Title: US-09-909-164-7
Perfect score: 56
Sequence: 1 EEVVPXGMHYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	73.2	413	11	Q8K289 mus musculus
2	39	69.6	1057	16	Q8CPJ4 staphylococ
3	37	66.1	308	16	Q9X2E2 thermotoga
4	37	66.1	322	17	Q9HLH8 oshl8 thermoplas
5	37	66.1	1044	16	Q8DIH0 oshl8 synecococc
6	36	64.3	208	2	Q46486 corynebacte
7	36	64.3	252	16	Q8YWP1 anabaena sp
8	36	64.3	819	10	Q9AVK4 pismu bacill
9	35	62.5	139	2	Q57489 bacteroides
10	35	62.5	156	16	Q9PC35 xyloella fas
11	35	62.5	233	10	Q40479 nicotiana t
12	35	62.5	237	10	Q9LW50 nicotiana s
13	35	62.5	317	9	Q38317 lactobacilli
14	35	62.5	425	5	Q9XVK4 caenorhabdi
15	35	62.5	510	10	Q9SA71 arabidopsis
16	35	62.5	2042	17	Q8T207 methanopyru

17	34	60.7	264	17	Q27902 methanobact
18	34	60.7	279	16	Q9RXN9 deinoxococcus
19	34	60.7	285	16	Q9EHU6 rhizobium 1
20	34	60.7	350	16	Q9RW92 deinoxococcus
21	34	60.7	355	16	Q9XAM3 streptomyces
22	34	60.7	360	17	Q27679 methanobact
23	34	60.7	425	16	Q9KBA1 bacillus ha
24	34	60.7	484	11	Q8CAL1 mus musculu
25	34	60.7	495	5	Q16912 caenorhabdi
26	34	60.7	637	5	Q9YOV6 ceratophila
27	34	60.7	678	12	Q9ELX6 ceratophila
28	34	60.7	748	12	Q9YR01 ranid herpe
29	34	60.7	992	11	Q8BNL0 mus musculu
30	34	60.7	1028	16	Q8YJ11 bruceella me
31	34	60.7	1070	11	Q8C6X2 mus musculu
32	34	60.7	2438	5	Q9VQJ7 drosophila
33	33	58.9	156	3	Q12479 saccharomyc
34	33	58.9	158	16	Q8PEP8 xanthomonas
35	33	58.9	208	2	Q8KTQ4 candidatus
36	33	58.9	216	16	Q9RDC1 streptomyces
37	33	58.9	254	16	Q8PIB0 xanthomonas
38	33	58.9	257	8	Q99011 prototheca
39	33	58.9	273	4	Q96N44 homo sapien
40	33	58.9	306	16	Q8XVQ7 ralsstonia s
41	33	58.9	364	5	Q8IGN5 drosophila
42	33	58.9	367	16	Q916J1 pseudomonas
43	33	58.9	387	16	Q96FX1 rhizobium 1
44	33	58.9	425	16	Q8DB81 vibrio vuln
45	33	58.9	426	3	Q9HG99 mortierella

ALIGNMENTS

RESULT 1
Q8K289 PRELIMINARY; PRT; 413 AA.
ID Q8K289
AC Q8K289;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Similar to expressed sequence A1987856.
GN 1110004B19RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Buthera; Rodentia; Sciurognath; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC032195; AAI32195.1; -;
DR MGD; MGI:1915724; 1110004B19RIK.
SQ SEQUENCE 413 AA; 46621 MW; A9B8A1DC70CDA0D5 CRC64;

Query Match 73.2%; Score 41; DB 11; Length 413;
Best Local Similarity 63.6%; Pred. No. 5.4;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EEVVPXGMHYS 11
Db 168 EEVVPXGMHYS 178
RESULT 2
Q8CPJ4 PRELIMINARY; PRT; 1057 AA.
ID Q8CPJ4
AC Q8CPJ4;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Carbamoyl-phosphate synthase large chain.
GN SE0879;

OS Straphylococcus epidermidis.
 OC Bacteria; Firmicutes; Bacillales; Straphylococcus.
 OX NCBI_TaxID=1282;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 12228;
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
 RL Chen Z., Wen Y.,
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB016746; AA004476.1; -
 KM Complete proteome.
 SQ SEQUENCE 1057 AA; 117391 MW; 8944D7D8DB1CA59 CRC64;

Query Match 69.6%; Score 39; DB 16; Length 1057;
 Best Local Similarity 63.6%; Pred. No. 38;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 1 EEVVPXGMHYS 11
 |||:|:|
 Db 189 KEVVSNGLHYS 199

RESULT 3

O9X2E2 PRELIMINARY; PRT; 308 AA.
 ID O9X2E2;
 AC O9X2E2;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE FTSH protease activity modulator HPLK.
 GN TM1822.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB / DSM 3109;
 RC MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 RA Halt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 RL Genome sequence of Thermotoga maritima.";
 RL Nature 399:323-329(1999).
 DR EMBL; AB001819; AAD36885.1; -
 DR TIGR; TM1822; -
 DR InterPro; IPR001107; Band 7;
 DR InterPro; IPR001972; Stomatin.
 DR Pfam; PF01145; Band 7; 1.
 DR PRINTS; PRO0721; STOMATIN.
 DR SMART; SMO0244; PHB. 1.
 KM Protease; Complete proteome.
 SQ SEQUENCE 308 AA; 34778 MW; ADE03603B5101A9D CRC64;

Query Match 66.1%; Score 37; DB 16; Length 308;
 Best Local Similarity 75.0%; Pred. No. 26;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 3 VVPXGMHYS 10
 |||:|:|
 Db 41 VVPSGIIHY 48

RESULT 4

O9HLH8 PRELIMINARY; PRT; 322 AA.
 ID O9HLH8;
 AC O9HLH8;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE Glucose-fructose oxidoreductase related protein.
 GN TM0250.
 OS Thermoplasma acidophilum.
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
 OC Archaea; Euryarchaeota; Thermoplasmata.
 OX NCBI_TaxID=2303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 1728;
 RX MEDLINE=20479972; PubMed=11029001;
 RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
 RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
 RT "The genome sequence of the thermophilic scavenger Thermoplasma
 RT acidophilum.";
 RL Nature 407:508-513(2000).
 DR EMBL; AL445063; CAC11395.1; -
 DR InterPro; IPR000683; GFO_IDH_MCCA.
 DR Pfam; PF01408; GFO_IDH_MCCA; 1.
 KM Complete proteome.
 SQ SEQUENCE 322 AA; 36918 MW; B8C239E71009D167 CRC64;

Query Match 66.1%; Score 37; DB 17; Length 322;
 Best Local Similarity 75.0%; Pred. No. 27;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 3 VVPXGMHYS 10
 |||:|:|
 Db 66 VVPDGLHY 73

RESULT 5

O8DIH0 PRELIMINARY; PRT; 1044 AA.
 ID O8DIH0;
 AC O8DIH0;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Multidrug efflux transporter.
 GN TLI1618.
 OS Synechococcus elongatus (Thermosynechococcus elongatus).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
 OX NCBI_TaxID=32046;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP-1;
 RC MEDLINE=2225144; PubMed=12240834;
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 RA Watanabe A., Iriyuchi M., Kawashima K., Kimura T., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A., Nakazaki N.,
 RA Shiino S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the thermophilic cyanobacterium
 RT Thermosynechococcus elongatus BP-1.";
 RL DNA Res. 9:123-130(2002).
 DR EMBL; AP005374; BAC09170.1; -
 KM Complete proteome.
 SQ SEQUENCE 1044 AA; 113205 MW; 00B9C13F0F636D2F CRC64;

Query Match 66.1%; Score 37; DB 16; Length 1044;
 Best Local Similarity 63.6%; Pred. No. 98;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 1 EEVVPXGMHYS 11
 |||:|:|
 Db 843 EEVLPNGIGYS 853

RESULT 6

O46486 PRELIMINARY; PRT; 208 AA.
 ID O46486;
 AC O46486;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)

DE Hypothetical 23.0 kDa protein (Gcra).
 GN Gcra.
 OS Corynebacterium xerosis, and
 OS Corynebacterium striatum.
 OC Bacteria; Actinobacteria; Actinomycetales;
 OC Corynebacteriaceae; Corynebacterium.
 OC NCBI_TaxID=1725, 43770;
 [1]
 RN
 RP
 RC
 RC SPECIES=C.xerosis; STRAIN=M82B;
 RX MEDLINE=96117603; PubMed=8559800;
 RA Tauch A., Kassing P., Kalinowski J., Puhler A.;
 RT "The Corynebacterium xerosis composite transposon Tns432 consists of
 RT two identical insertion sequences, designated IS1249, flanking the
 RT erythromycin resistance gene ermC.",
 RL Plasmid 34.113-131 (1995).
 [2]
 RN
 RP
 RC
 RC SPECIES=C.striatum; STRAIN=M82B;
 RX MEDLINE=20194806; PubMed=10732668;
 RA Tauch A., Krieft S., Kalinowski J., Puhler A.;
 RT "The 51,403-bp R-plasmid pRP10 from the multiresistant clinical
 RT isolate Corynebacterium striatum M82B is composed of DNA segments
 RT initially identified in soil bacteria and in plant, animal, and human
 RT pathogens".
 RL Mol. Gen. Genet. 263:1-11(2000).
 DR EMBL; U21300; AAC95478.1; -.
 DR EMBL; AF024666; AAG03390.1; -.
 KW Hypothetical protein; Plasmid.
 SQ SEQUENCE 208 AA; 23012 MW; F1504B1ECDE85A6 CRC64;

Query Match 64.3%; Score 36; DB 2; Length 208;
 Best Local Similarity 50.0%; Pred. No. 27;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVFXGMHY 11
 :|||:
 Db 130 DVIPEGKHYA 139

RESULT 7
 O8YWP1 PRELIMINARY; PRT; 252 AA.
 AC O8YWP1;
 DT 01-MAR-2002 (TRENBLREL. 20, Created)
 DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)
 DT 01-MAR-2002 (TRENBLREL. 20, Last annotation update)
 DE Hypothetical protein AL1563.
 GN AL1563.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OC NCBI_TaxID=103690;
 [1]
 RN
 RP
 RC
 RC SPECIES=FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Kawanabe A., Iriuch N., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabaata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 RT cyanobacterium Anabaena sp. strain PCC 7120."
 RL CNA Res. 8:205-213(2001).
 DR EMBL; AP003586; BAB77929.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 252 AA; 28831 MW; 925572DASD1CA519 CRC64;

Query Match 64.3%; Score 36; DB 16; Length 252;
 Best Local Similarity 50.0%; Pred. No. 34;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 EVVFXGMHY 10
 :|||:
 :|||:

Db 235 EMIVPAGLHF 244

RESULT 8
 O9AVK4 PRELIMINARY; PRT; 819 AA.
 AC O9AVK4;
 DT 01-JUN-2001 (TRENBLREL. 17, Created)
 DT 01-JUN-2001 (TRENBLREL. 17, Last sequence update)
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE SCARECROW.
 GN PSSCR.
 OS Pisum sativum (Garden pea).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eubotids 1; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
 OC NCBI_TaxID=3888;
 [1]
 RN
 RP
 RC
 RC SPECIES=FROM N.A.
 RX STRAIN=cv. Alaska;
 RX MEDLINE=21231727; PubMed=11333309;
 RA Sadaa N., Matsushita Y., Nakamura T., Nymoya H.;
 RT "The Molecular Characterization and in situ Expression Pattern of Pea
 RT SCARECROW Gene."
 RL Plant Cell Physiol. 42:385-394(2001).
 DR EMBL; AB048713; BAB39155.1; -.
 DR InterPro; IPR001444; Pfam_Pfam_rod.
 DR InterPro; IPR005202; GRAS.
 DR Pfam; PF03514; GRAS; 1.
 DR PROSITE; PS00588; FLAGELLA_BB_ROD; 1.
 SQ SEQUENCE 819 AA; 90372 MW; 41B67BD6DC72ADFA CRC64;

Query Match 64.3%; Score 36; DB 10; Length 819;
 Best Local Similarity 45.5%; Pred. No. 1.2e+02;
 Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EVVFXGMHY 11
 :|||:
 Db 343 DDVPTSLHFS 353

RESULT 9
 O57489 PRELIMINARY; PRT; 139 AA.
 AC O57489;
 DT 01-NOV-1996 (TRENBLREL. 01, Created)
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
 DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
 DE DNA ligase (fragment)
 GN Bacteroides nodosus (Dichelobacter nodosus).
 OS Bacteroides nodosus (Dichelobacter nodosus).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Cardiobacteriales;
 OC Cardiobacteriaceae; Dichelobacter.
 OC NCBI_TaxID=870;
 [1]
 RN
 RP
 RC
 RC SPECIES=FROM N.A.
 RX MEDLINE=96020672; PubMed=7476204;
 RA Moses E.K., Good R.T., Sinistat M., Billington S.J., Langford C.J.,
 RA Wood J.I.;
 RT "A multiple site-specific DNA-inversion model for the control of Omp1
 RT phase and antigenic variation in Dichelobacter nodosus."
 RL Mol. Microbiol. 17:183-196(1995).
 [2]
 RN
 RP
 RC
 RC SPECIES=FROM N.A.
 RX MEDLINE=96257263; PubMed=8654969;
 RA Billington S.J., Sinistat M., Cheetham B.F., Ayres A., Moses E.K.,
 RA Katz M.E., Wood J.I.;
 RT "Identification of a native Dichelobacter nodosus plasmid and
 RT implications for the evolution of the vap regions."
 RL Gene 172:111-116(1996).
 DR EMBL; U02462; AAB12366.1; -.
 DR InterPro; IPR001357; BRCT.
 DR Pfam; PF00533; BRCT; 1.
 DR SMART; SM00292; BRCT; 1.

DR PROSITE: PS50172; BRCT; 1.
 KW Ligase.
 FT NOY TER
 SQ SEQUENCE 139 AA; 15052 MW; E0E110AA4B7D4708 CRC64;

Query Match 62.5%; Score 35; DB 2; Length 139;
 Best Local Similarity 55.6%; Pred. No. 28;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 3 VVVPXGMH 11
 :|||:|:
 Db 21 IVPAGVHMS 29

RESULT 10

O9PC35 PRELIMINARY; PRT; 156 AA.
 ID O9PC35
 AC O9PC35;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Hypothetical protein Xf1950.
 GN Xf1950.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Artuda P., Abreu F.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.B.A., Carraro D.M., Carter H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohenberg J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite J.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.V., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,
 RA Marques C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E.Jr., de Sa R.G., Sanchelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A.Jr.,
 RA de Souza A.P., Terezi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidants J., Setubal J.C., "The genome sequence of the plant pathogen Xylella fastidiosa.";
 RL Nature 406:151-159(2000).
 DR EMBL; AE004014; AAF84752.1;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 156 AA; 17144 MW; D8558619C6671A5D CRC64;

Query Match 62.5%; Score 35; DB 16; Length 156;
 Best Local Similarity 55.6%; Pred. No. 32;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 EEVVPXGMH 9
 :|||:|:
 Db 119 EEIIPGVH 127

RESULT 11

O40479 PRELIMINARY; PRT; 233 AA.
 ID O40479
 AC O40479;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE ERBP-2.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BY4; TISSUE=leaf;
 RX MEDLINE=95276459; PubMed=7756828;
 RA Ohme-Takagi M., Shinshi H.,
 RT "Ethylene-inducible DNA binding proteins that interact with an
 RT ethylene responsive element.";
 RL Plant Cell 7:173-182(1995).
 DR EMBL; D38126; BAA07324.1;
 DR HSSP; O80337; 2GCC.
 DR TRANSFAC; T02654;
 DR InterPro; IPR001471; TF_ERF.
 DR Pfam; PF00847; AP2-domain; 1.
 DR ProDom; PD001423; TF_ERF; 1.
 DR SMART; SM00380; AP2_1.
 SQ SEQUENCE 233 AA; 25563 MW; 6CD16783582C0CB5 CRC64;

Query Match 62.5%; Score 35; DB 10; Length 233;
 Best Local Similarity 60.0%; Pred. No. 50;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 EEVVPXGMH 10
 :|||:|:
 Db 90 QAVVPKGRHY 99

RESULT 12

O9LW50 PRELIMINARY; PRT; 237 AA.
 ID O9LW50
 AC O9LW50;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Ethylene-responsive element binding factor.
 GN NERF2.
 OS Nicotiana glauca (Wood tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4096;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20399450; PubMed=10945353;
 RX Kitajima S., Koyama T., Ohme-Takagi M., Shinshi H., Sato F.;
 RT "Characterization of gene expression of NERF2, transcription factors
 RT of basic PR genes from Nicotiana glauca.";
 RL Plant Cell Physiol. 41:817-824(2000).
 DR EMBL; AB016264; BAA97122.1;
 DR HSSP; O80337; 2GCC.
 DR InterPro; IPR001471; TF_ERF.
 DR Pfam; PF00847; AP2-domain; 1.
 DR ProDom; PD001423; TF_ERF; 1.
 DR SMART; SM00380; AP2_1.
 SQ SEQUENCE 237 AA; 26243 MW; 01BC3EB51E46298 CRC64;

Query Match 62.5%; Score 35; DB 10; Length 237;
 Best Local Similarity 60.0%; Pred. No. 51;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 EEVVPXGMH 10
 :|||:|:

Db 94 QAVVPGKMHY 103

RESULT 13

Q38317 PRELIMINARY; PRT; 317 AA.

AC Q38317; (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE Lysin.

GN Lys.

OS Lactobacillus bacteriophage phi adh.

OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.

OX NCBI_TaxID=12417;

RN [1] Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.

RA Altermann E.;

RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99384014; PubMed=10452953;

RA Altermann E., Klein J., Henrich B.;

RT "Primary structure and features of the genome of the Lactobacillus gasseri temperate bacteriophage phi-adh.";

RL Gene 236:333-346(1999).

RL [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=95138034; PubMed=7836307;

RA Henrich B., Binsthofer B., Blesi U.;

RT "Primary structure and functional analysis of the lysis genes of Lactobacillus gasseri bacteriophage phi-adh.";

RL J. Bacteriol. 177:723-732(1995).

RN [4]

RP SEQUENCE FROM N.A.

RX MEDLINE=93231538; PubMed=8472961;

RA Fremaux C., De Antoni G., Raya R., Klamhammer T.;

RT "Genetic organization and sequence of the region encoding integrative functions from Lactobacillus gasseri temperate bacteriophage phi-adh.";

RL Gene 126:61-66(1993).

RN [5]

RP SEQUENCE FROM N.A.

RA Engel G., Altermann E., Klein J., Henrich B.;

RT "Structure of a genome region of the Lactobacillus gasseri temperate phage phi adh covering a repressor gene and cognate promoters.";

RL Gene 210:67-70(1998).

DR EMBL; AJ11519; CAB52540.1; --

DR InterPro; IPR002053; Glyco_hydro_25.

DR Pfam; PF01183; Glyco_hydro_25; 1.

DR ProDom; PD004620; Glyco_hydro_25; 1.

DR SMART; SMO0641; Glyco_25; 1.

DR SMART; SMO0287; SH3b; 1.

SO SEQUENCE 317 AA; 34703 MW; 9FF2715EE43561C7 CRC64;

Qy 3 VVPXGMHYS 11

Db 60 VVPMGTHYA 68

Query Match 62.5%; Score 35; DB 9; Length 317;

Best Local Similarity 66.7%; Pred. No. 69;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

RESULT 14

Q9XVKA PRELIMINARY; PRT; 425 AA.

DT 01-NOV-1999 (TREMBlrel. 12, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE R10D12.10 protein.

GN R10D12.10.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA Percy C.M.;

RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA none;

RT "Genome sequence of the nematode C.elegans: A platform for investigating biology.";

RL Science 282:2012-2018(1998).

DR EMBL; Z81109; CAB03241.1; --

DR WormPep; R10D12.10; CE12690.

DR InterPro; IPR000719; Prot_kinase.

DR Pfam; PF00069; kinase; 1.

DR ProDom; PD000001; Prot_kinase; 1.

DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.

KW ATP-binding; Transferase.

SO SEQUENCE 425 AA; 49410 MW; 5D96E29B08C8E9D6 CRC64;

Query Match 62.5%; Score 35; DB 5; Length 425;

Best Local Similarity 50.0%; Pred. No. 95;

Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EEVVPXGMH 10

Db 335 EQVPGSLQY 344

RESULT 15

Q9SA71 PRELIMINARY; PRT; 510 AA.

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE T518.1 protein.

GN T518.1.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STAIN=cv. Columbia.

RA Vysotskaia V.S., Schwartz J.R., Yu G., Tortum M., Lenz C., Liu S., Li J., Kremetsekaia I., Luros J., Ngan I., Gonzalez A., Altati H., Araujo R., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N., Huitzer L., Kim C., Palm C., Rowley D., Shinn P., Walker M., Davis R.W., Becker J.R., Federspiel N.A., Theologis A.;

RT "Arabidopsis thaliana chromosome 1 BAC T518 sequence.";

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC007060; AAD25743.1; --

DR InterPro; IPR007114; MFS.

DR PROSITE; PSS0850; MFS; 1.

SO SEQUENCE 510 AA; 55089 MW; 0DC63CC41F4712F8 CRC64;

Qy 1 EEVVPXGMH 10

Db 12 EEVXPPGIR 21

Query Match 62.5%; Score 35; DB 10; Length 510;

Best Local Similarity 60.0%; Pred. No. 1,2e+02;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Job time : 27.2667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using BW model

Run on: December 22, 2003, 16:41:00 ; Search time 32.4667 Seconds
(without alignments)
53.778 Million cell updates/sec

Title: US-09-909-164-8

Sequence: 1 EEVVPXGMDYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	96.3	11	23	ABBB0524 Hepatitis C virus
2	52	96.3	11	23	ABBB0528 Hepatitis C virus
3	52	96.3	11	23	ABBB0529 Hepatitis C virus
4	52	96.3	11	23	ABBB0561 Hepatitis C virus
5	52	96.3	11	23	ABBB0562 Hepatitis C virus
6	47	87.0	11	23	ABBB0538 Hepatitis C virus
7	47	87.0	11	23	ABBB0542 Hepatitis C virus
8	47	87.0	11	23	ABBB0543 Hepatitis C virus
9	46	85.2	11	23	ABBB0521 Hepatitis C virus

10	46	85.2	11	23	ABBB0522 Hepatitis C virus
11	46	85.2	11	23	ABBB0525 Hepatitis C virus
12	46	85.2	11	23	ABBB0526 Hepatitis C virus
13	46	85.2	11	23	ABBB0547 Hepatitis C virus
14	46	85.2	11	23	ABBB0548 Hepatitis C virus
15	46	85.2	11	23	ABBB0551 Hepatitis C virus
16	46	85.2	11	23	ABBB0556 Hepatitis C virus
17	46	85.2	11	23	ABBB0557 Hepatitis C virus
18	46	85.2	11	23	ABBB0559 Hepatitis C virus
19	46	85.2	11	23	ABBB0563 Hepatitis C virus
20	46	85.2	11	23	ABBB0564 Hepatitis C virus
21	46	85.2	11	23	ABBB0565 Hepatitis C virus
22	46	85.2	11	23	ABBB0567 Hepatitis C virus
23	46	85.2	11	23	ABBB0568 Hepatitis C virus
24	46	85.2	11	23	ABBB0568 Hepatitis C virus
25	45	83.3	11	23	ABBB0523 Hepatitis C virus
26	45	83.3	11	23	ABBB0527 Hepatitis C virus
27	45	83.3	11	23	ABBB0558 Hepatitis C virus
28	45	83.3	11	23	ABBB0560 Hepatitis C virus
29	44	81.5	11	23	ABBB0533 Hepatitis C virus
30	44	81.5	11	23	ABBB0534 Hepatitis C virus
31	41	75.9	11	23	ABBB0535 Hepatitis C virus
32	41	75.9	11	23	ABBB0536 Hepatitis C virus
33	41	75.9	11	23	ABBB0539 Hepatitis C virus
34	41	75.9	11	23	ABBB0540 Hepatitis C virus
35	40	74.1	11	23	ABBB0537 Hepatitis C virus
36	40	74.1	11	23	ABBB0541 Hepatitis C virus
37	40	74.1	11	23	ABBB0544 Hepatitis C virus
38	40	74.1	11	23	ABBB0545 Hepatitis C virus
39	40	74.1	11	23	ABBB0549 Hepatitis C virus
40	40	74.1	11	23	ABBB0552 Hepatitis C virus
41	40	74.1	11	23	ABBB0553 Hepatitis C virus
42	39	72.2	11	23	ABBB0546 Hepatitis C virus
43	39	72.2	11	23	ABBB0550 Hepatitis C virus
44	39	72.2	11	23	ABBB0554 Hepatitis C virus
45	39	72.2	11	23	ABBB0555 Hepatitis C virus

ALIGNMENTS

RESULT 1	ABBB0524	ID	ABBB0524	standard; peptide; 11 AA.
XX	AC	ABBB0524;		
XX	DT	08-OCT-2002	(first entry)	
XX	DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.		
XX	KW	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.		
XX	OS	Synthetic.		
XX	FT	Key	Location/Qualifiers	
FT	FT	Modified-site 1	/note= "N-terminal acetyl"	
FT	FT	Modified-site 6	/note= "N-terminal acetyl"	
FT	FT	Modified-site 9	/note= "N-terminal acetyl"	
FT	FT	Misc-difference 9	/note= "D-form residue"	
FT	FT	Modified-site 11	/note= "C-terminal amide"	
PN	PN	W0200208251-A2.		
XX	PD	31-JAN-2002.		
XX	PF	19-JUL-2001; 2001WO-US23169.		
XX	XX			

PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Llm-wilby M, Levy OE, Brunck TK;
XX
DR WPI, 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
PS Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

SQ Sequence 11 AA;

Query Match 96.3%; Score 52; DB 23; Length 11;

Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EEVVPXGMDYS 11
|||
1 EEVVPXGMDYS 11

Db

RESULT 2
ABB80528
ID ABB80528 standard; peptide; 11 AA.

XX ABB80528;

XX 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #8.

XX Hepatitis C virus; HCV, serine protease; inhibitor; alpha-ketoamide;
XX virucide.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "N-terminyl carbonyl forming keto-amide linkage with
FT residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

XX W0200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Llm-wilby M, Levy OE, Brunck TK;

XX WPI, 2002-361643/39.

PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
PS Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

SQ Sequence 11 AA;

Query Match 96.3%; Score 52; DB 23; Length 11;

Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EEVVPXGMDYS 11
|||
1 EEVVPXGMDYS 11

Db

RESULT 3
ABB80529
ID ABB80529 standard; peptide; 11 AA.

XX ABB80529;

XX 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.

XX Hepatitis C virus; HCV, serine protease; inhibitor; alpha-ketoamide;
XX virucide.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "N-terminyl carbonyl forming keto-amide linkage with
FT residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

XX W0200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Llm-wilby M, Levy OE, Brunck TK;

XX WPI, 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
PS Claim 17; Page 64; 69pp; English.

Query Match 96.3%; Score 52; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
DB 1 EEVVPXGMDYS 11

RESULT 6
ABB80538
ID ABB80538 standard; peptide; 11 AA.

AC ABB80538;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #18.

KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
virucide.

OS Synthetic.

Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "N-terminal carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

PN WO200208251-A2.

PD 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

PA (CORV-) CORVAS INT INC.

PI Lim-wilby M, Levy OE, Brunck TK;

DR WPI; 2002-361643/39.

PT Novel peptide compound having hepatitis C virus protease inhibitory
activity useful for treating disorders associated with hepatitis C
virus protease -
PS Claim 17; Page 64; 69pp; English.

CC The sequence represents a peptide compound of the invention having
hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
invention are alpha-ketoamide peptide analogues. The peptides have
virucide activity, and are useful for treating and in the manufacture of
a medicament to treat disorders associated with HCV protease. A
pharmaceutical composition comprising the peptide as an active ingredient
is useful for treating disorders associated with hepatitis C virus.

CC Sequence 11 AA;

Query Match 87.0%; Score 47; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.017;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
DB 1 EEVVPXGMDYS 11

RESULT 7
ABB80542
ID ABB80542 standard; peptide; 11 AA.

AC ABB80542;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #22.

KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
virucide.

OS Synthetic.

Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "N-terminal carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

PN WO200208251-A2.

PD 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

PA (CORV-) CORVAS INT INC.

PI Lim-wilby M, Levy OE, Brunck TK;

DR WPI; 2002-361643/39.

PT Novel peptide compound having hepatitis C virus protease inhibitory
activity useful for treating disorders associated with hepatitis C
virus protease -
PS Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having
hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
invention are alpha-ketoamide peptide analogues. The peptides have
virucide activity, and are useful for treating and in the manufacture of
a medicament to treat disorders associated with HCV protease. A
pharmaceutical composition comprising the peptide as an active ingredient
is useful for treating disorders associated with hepatitis C virus.

CC Sequence 11 AA;

Query Match 87.0%; Score 47; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.017;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
DB 1 EEVVPXGMDYS 11

RESULT 8

ABB80543
ID ABB80543 standard; peptide; 11 AA.

AC ABB80543;

DT 08-OCT-2002 (first entry)

```

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #23.
DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
KM
XX Synthetic.
OS
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT Modified-site /note= residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
XX WO200208251-A2.
PN 31-JAN-2002.
PD
XX 19-JUL-2001; 2001WO-US23169.
PE
XX 21-JUL-2000; 2000US-220101P.
PR
XX (CORV-) CORVAS INT INC.
PA
XX Lim-wilby M, Levy OE, Brunck TK;
PI
XX WPI; 2002-361643/39.
DR
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 65; 69pp; English.
PS
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
CC
XX Sequence 11 AA;
SQ
Query Match 87.0%; Score 47; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.017;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 BEVVPXGMDYS 11
DB 1 BEVVPXGMDYS 11

```

```

OS Synthetic.
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT Modified-site /note= residue 7"
FT Modified-site 11 /note= "C-terminal amide"
FT
XX WO200208251-A2.
PN 31-JAN-2002.
PD
XX 19-JUL-2001; 2001WO-US23169.
PE
XX 21-JUL-2000; 2000US-220101P.
PR
XX (CORV-) CORVAS INT INC.
PA
XX Lim-wilby M, Levy OE, Brunck TK;
PI
XX WPI; 2002-361643/39.
DR
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
PS
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
CC
XX Sequence 11 AA;
SQ
Query Match 85.2%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.026;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 BEVVPXGMDYS 11
DB 1 BEVVPXGMDYS 11

```

```

RESULT 9
ABB80521
ID ABB80521 standard; peptide; 11 AA.
AC ABB80521;
XX
XX 08-OCT-2002 (first entry)
DT
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.
DE
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM virucide.
XX

```

```

RESULT 10
ABB80522
ID ABB80522 standard; peptide; 11 AA.
AC ABB80522;
XX
XX 08-OCT-2002 (first entry)
DT
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.
DE
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM virucide.
XX
XX Synthetic.
OS
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT Modified-site /note= residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT

```

FT Modified-site 11 /note= "C-terminal amide"
XX
PN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
PS Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 85.2%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.026;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMDYS 11
DB 1 EEVVPXGMSYS 11
RESULT 11
ABB80525
ID ABB80525 standard; peptide; 11 AA.
XX
AC ABB80525;
XX
XX 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #5.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
OS Synthetic.
XX
XX Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.

XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
PS Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 85.2%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.026;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMDYS 11
DB 1 EEVVPXGMSYS 11
RESULT 12
ABB80526
ID ABB80526 standard; peptide; 11 AA.
XX
AC ABB80526;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #6.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
OS Synthetic.
XX
XX Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX

Query Match 85.2%; Score 46; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.026; 1; Indels 0; Gaps 0;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Search completed: December 22, 2003, 17:41:00
 Job time : 32.4667 secs

QY 1 EEVVPXGMDYS 11
 |||||
 DB 1 EEVVPXGTDYS 11

RESULT 15

ABB80551
 ID ABB80551 standard; peptide; 11 AA.

AC ABB80551;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #31.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.

OS Synthetic.

PH Key Location/Qualifiers

FT Modified-site 1

FT /note= "N-terminal acetyl"

FT Modified-site 6

FT /note= "N-terminal carbonyl forming keto-amide linkage with
 residue 7"

FT Misc-difference 9

FT /note= "D-form residue"

FT Modified-site 11

FT /note= "C-terminal amide"

PN WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Llm-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -

XX Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 CC

SQ Sequence 11 AA;

Query Match 85.2%; Score 46; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.026; 1; Indels 0; Gaps 0;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
 |||||
 DB 1 EEVVPXGSDYS 11

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 16:37:03 / Search time 10.1333 Seconds
(without alignments)
45.930 Million cell updates/sec

Title: US-09-909-164-8

Perfect score: 54

Sequence: 1 EEVVPXGMDYS 11

Scoring table: BIOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	36	66.7	341	3	US-08-853-948B-4
2	36	66.7	348	3	US-08-853-948B-5
3	36	66.7	368	4	US-09-697-367-24
4	36	66.7	1049	4	US-09-394-272-10
5	36	66.7	1068	2	US-08-429-054A-11
6	36	66.7	1068	2	US-08-718-777-7
7	36	66.7	1068	3	US-09-051-341-7
8	36	66.7	1068	4	US-09-394-272-8
9	36	66.7	1081	4	US-09-394-272-4
10	36	66.7	1083	4	US-09-394-272-11
11	36	66.7	1084	4	US-09-394-272-9
12	36	63.0	140	3	US-08-569-147-76
13	34	63.0	140	3	US-08-569-147-82
14	34	63.0	1065	4	US-09-252-991A-21637
15	33	61.1	59	4	US-08-963-851-14
16	33	61.1	378	1	US-08-070-165F-8
17	33	61.1	378	2	US-08-888-418-8
18	33	61.1	473	4	US-09-252-991A-26805
19	33	61.1	765	4	US-09-819-989-4
20	33	61.1	801	3	US-09-383-630-6
21	33	61.1	811	4	US-09-819-989-2
22	33	61.1	883	4	US-09-667-373-4
23	32	59.3	65	6	5177197-51
24	32	59.3	102	2	US-08-580-968A-23
25	32	59.3	152	2	US-08-460-694-4
26	32	59.3	152	3	US-08-460-744-4
27	32	59.3	152	3	US-07-667-711B-4

28	32	59.3	173	1	US-08-193-977-7	Sequence 7, Appli
29	32	59.3	189	2	US-08-464-517-21	Sequence 21, Appli
30	32	59.3	189	2	US-08-246-361A-21	Sequence 21, Appli
31	32	59.3	189	3	US-08-463-772-21	Sequence 21, Appli
32	32	59.3	189	5	PCT-US93-05000-21	Sequence 21, Appli
33	32	59.3	231	3	US-08-926-842B-20	Sequence 20, Appli
34	32	59.3	236	2	US-08-464-517-22	Sequence 22, Appli
35	32	59.3	236	2	US-08-246-361A-22	Sequence 22, Appli
36	32	59.3	236	3	US-08-463-772-22	Sequence 22, Appli
37	32	59.3	236	5	PCT-US93-05000-22	Sequence 22, Appli
38	32	59.3	240	3	US-08-926-842B-21	Sequence 21, Appli
39	32	59.3	241	4	US-09-107-532A-4172	Sequence 4172, Ap
40	32	59.3	280	2	US-08-464-517-6	Sequence 6, Appli
41	32	59.3	280	3	US-08-463-772-6	Sequence 6, Appli
42	32	59.3	289	2	US-08-246-361A-4	Sequence 4, Appli
43	32	59.3	289	5	PCT-US93-05000-4	Sequence 4, Appli
44	32	59.3	291	5	PCT-US93-05000-6	Sequence 6, Appli
45	32	59.3	292	2	US-08-464-517-23	Sequence 23, Appli

ALIGNMENTS

RESULT 1
US-08-853-948B-4
; Sequence 4, Application US/08853948B
; Patent No. 6210943
; GENERAL INFORMATION:
; APPLICANT: AKIHAMA, Toyota
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE FROM CITRUS AND DNA ENCODING
; FILE REFERENCE: 0049-0235-0
; CURRENT APPLICATION NUMBER: US/08/853,948B
; CURRENT FILING DATE: 1997-05-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Citrus unshiu
; FEATURE:
; OTHER INFORMATION: Xaa at position 109 is one of Ala, Arg, Asn, Asp,
; OTHER INFORMATION: Cys, Gln, Gly, His, Ile, Leu, Lys, Met, Phe,
; OTHER INFORMATION: Pro, Ser, Thr, Trp, Tyr, or Val
US-08-853-948B-4

Query Match 66.7%; Score 36; DB 3; Length 341;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

CY 3 VVPPXGMDYS 11
DB 228 VVPPXGMDYS 236

RESULT 2
US-08-853-948B-5
; Sequence 5, Application US/08853948B
; Patent No. 6210943
; GENERAL INFORMATION:
; APPLICANT: AKIHAMA, Toyota
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE FROM CITRUS AND DNA ENCODING
; FILE REFERENCE: 0049-0235-0
; CURRENT APPLICATION NUMBER: US/08/853,948B
; CURRENT FILING DATE: 1997-05-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Citrus unshiu
US-08-853-948B-5

Query Match 66.7%; Score 36; DB 3; Length 348;
Best Local Similarity 66.7%; Pred. No. 19;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 VVPXGMDYS 11
|:| |:| |:|
Db 234 VVPGMDPS 242

RESULT 3

US-09-697-367-24
; Sequence 24, Application US/09697367
; Patent No. 6323015
; GENERAL INFORMATION:
; APPLICANT: Orozco Jr., Emil M.
; APPLICANT: Caimi, Perry G.
; APPLICANT: Meng, Zude
; APPLICANT: Tarczynski, Mitchell
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: B8116 US NA
; CURRENT FILING DATE: US/09/697,367
; PRIOR FILING DATE: 2000-10-26
; PRIOR APPLICATION NUMBER: 60/084,529
; PRIOR FILING DATE: 1998-MAY-07
; PRIOR APPLICATION NUMBER: PCT/US99/09865
; PRIOR FILING DATE: 1999-MAY-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Zea mays
US-09-697-367-24

Query Match 66.7%; Score 36; DB 4; Length 368;
Best Local Similarity 66.7%; Pred. No. 20;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 VVPXGMDYS 11
|:| |:| |:|
Db 217 VVPGMDPS 225

RESULT 4

US-09-394-272-10
; Sequence 10, Application US/09394272
; Patent No. 6472588
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; FILE REFERENCE: 201304/1000
; CURRENT FILING DATE: US/09/394,272
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1049
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-394-272-10

Query Match 66.7%; Score 36; DB 4; Length 1049;
Best Local Similarity 66.7%; Pred. No. 67;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 VVPXGMDYS 11
|:| |:| |:|
Db 436 VVPGMDPS 444

RESULT 5

US-08-429-054A-11
; Sequence 11, Application US/08429054A
; Patent No. 5917126
; GENERAL INFORMATION:
; APPLICANT: VAN ASSCHE, CHARLES; LANDO, DANIELLE; BRUNEAU,
; APPLICANT: JEAN; VOELKER, TONI; GERVAIS, MONICA
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHETASE (SPS),
; TITLE OF INVENTION: A PREPARATION METHOD AND CDNA THEREFOR, AND USE OF THE
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN AND MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/429,054A
; FILING DATE: 26-APR-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 842,337
; FILING DATE: 20-March-1992
; APPLICATION NUMBER: PCT/FR 91/00593
; FILING DATE: 18-July-1991
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: French 90402094.9
; FILING DATE: 20-July-1990
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles A. Musetlian
; REGISTRATION NUMBER: 19,683
; REFERENCE/DOCKET NUMBER: 146,1137
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 661-8000
; TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULAR TYPE: Peptide
US-08-429-054A-11

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 68;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 VVPXGMDYS 11
|:| |:| |:|
Db 435 VVPGMDPS 443

RESULT 6

US-08-718-777-7
; Sequence 7, Application US/08718777
; Patent No. 5981852
; GENERAL INFORMATION:
; APPLICANT: Van Assche, C.
; APPLICANT: Lando, D.
; APPLICANT: Bruneau, J. M.
; APPLICANT: Voelker, T.
; APPLICANT: Gervais, M.
; TITLE OF INVENTION: MODIFICATION OF SUCROSE

TITLE OF INVENTION: PHOSPHATE
TITLE OF INVENTION: SYNTHASE IN PLANTS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Barbara Rae-Venter
STREET: 260 Sheridan Avenue, Suite 440
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/718,777
FILING DATE: NOT YET ASSIGNED
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,471
FILING DATE: 27-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Barbara Rae-Venter
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: CGNE.072.02US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)328-4400
TELEFAX: (415)328-4477
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1068 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-718-777-7
Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. NO. 68;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Oy 3 VVPGXMDYS 11
|:| |||:|
Db 435 VVPGMDPS 443
RESULT 7
US-09-051-341-7
Sequence 7, Application US/09051341
Patent No. 6124528
GENERAL INFORMATION:
APPLICANT: Shewmaker, C. K.
TITLE OF INVENTION: MODIFICATION OF SOLUBLE SOLIDS USING
TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE ENCODING SEQUENCE
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rae-Venter Law Group, P.C.
STREET: 260 Sheridan Avenue, Suite 440
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/051,341
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/17351

FILING DATE: 25-OCT-1996
APPLICATION NUMBER: US 08/549,016
FILING DATE: 27-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/372,200
FILING DATE: 12-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Barbara Rae-Venter, Ph.D.,
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: CGNE.110.02US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)328-4400
TELEFAX: (415)328-4477
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1068 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-051-341-7
Query Match 66.7%; Score 36; DB 3; Length 1068;
Best Local Similarity 66.7%; Pred. NO. 68;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Oy 3 VVPGXMDYS 11
|:| |||:|
Db 435 VVPGMDPS 443
RESULT 8
US-09-394-272-8
Sequence 8, Application US/09394272
Patent No. 6472588
GENERAL INFORMATION:
APPLICANT: Haigler, Candace H.
TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
FILE REFERENCE: 201304/1000
CURRENT APPLICATION NUMBER: US/09/394,272
CURRENT FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 8
LENGTH: 1068
TYPE: PRT
ORGANISM: Zea mays
US-09-394-272-8
Query Match 66.7%; Score 36; DB 4; Length 1068;
Best Local Similarity 66.7%; Pred. NO. 68;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Oy 3 VVPGXMDYS 11
|:| |||:|
Db 435 VVPGMDPS 443
RESULT 9
US-09-394-272-4
Sequence 4, Application US/09394272
Patent No. 6472588
GENERAL INFORMATION:
APPLICANT: Haigler, Candace H.
TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
FILE REFERENCE: 201304/1000
CURRENT APPLICATION NUMBER: US/09/394,272
CURRENT FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 4
LENGTH: 1081
TYPE: PRT
ORGANISM: Craterostigma plantagineum
US-09-394-272-4

Query Match 66.7% Score 36; DB 4; Length 1081;
Best Local Similarity 66.7% Pred. No. 69;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
DB 445 VVPGMDPS 453

RESULT 10
US-09-394-272-11
Sequence 11, Application US/09394272
Patent No. 6472588
GENERAL INFORMATION:
APPLICANT: Haigler, Candace H.
APPLICANT: Haigler, A. Scott
TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
FILE REFERENCE: 201304/1000
CURRENT APPLICATION NUMBER: US/09/394,272
CURRENT FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 11
LENGTH: 1083
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-09-394-272-11

Query Match 66.7% Score 36; DB 4; Length 1083;
Best Local Similarity 66.7% Pred. No. 69;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
DB 483 VVPGMDPS 491

RESULT 11
US-09-394-272-9
Sequence 9, Application US/09394272
Patent No. 6472588
GENERAL INFORMATION:
APPLICANT: Haigler, Candace H.
APPLICANT: Haigler, A. Scott
TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
FILE REFERENCE: 201304/1000
CURRENT APPLICATION NUMBER: US/09/394,272
CURRENT FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 9
LENGTH: 1084
TYPE: PRT
ORGANISM: Oryza sativa
US-09-394-272-9

Query Match 66.7% Score 36; DB 4; Length 1084;
Best Local Similarity 66.7% Pred. No. 69;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
DB 453 VVPGMDPS 461

RESULT 12
US-08-569-147-76
Sequence 76, Application US/08569147
Patent No. 6180377
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: HUMANISED ANTIBODIES
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
ADDRESSEE: No. 61803771s, LLP
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: U.S.A.
ZIP: 19103

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/569,147
FILING DATE: 25-March-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Trujillo, Doreen Yanko
REGISTRATION NUMBER: 35,719
REFERENCE/DOCKET NUMBER: CARP-0047
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439

INFORMATION FOR SEQ ID NO: 76:
SEQUENCE CHARACTERISTICS:
LENGTH: 140 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-569-147-76

Query Match 63.0% Score 34; DB 3; Length 140;
Best Local Similarity 75.0% Pred. No. 16;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGMDY 10
DB 122 VVPGGPDY 129

RESULT 13
US-08-569-147-82
Sequence 82, Application US/08569147
Patent No. 6180377
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: HUMANISED ANTIBODIES
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
ADDRESSEE: No. 61803771s, LLP
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: U.S.A.
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/569,147
FILING DATE: 25-March-1996

```
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Trujillo, Doreen Yanko
; REGISTRATION NUMBER: 35,719
; REFERENCE/DOCKET NUMBER: CARP-0047
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-569-147-82

Query Match          63.0%; Score 34; DB 3; Length 140;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 VVFXGMDY 10
Db      122 VVPTGFDY 129

RESULT 14
US-09-252-991A-31637
; Sequence 31637, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31637
; LENGTH: 1065
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-31637

Query Match          63.0%; Score 34; DB 4; Length 1065;
Best Local Similarity 85.7%; Pred. No. 1,7e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      5 PYGMDYS 11
Db      324 PQGMDYS 330

RESULT 15
US-08-963-851-14
; Sequence 14, Application US/08963851
; Patent No. 6300116
; GENERAL INFORMATION:
; APPLICANT: VAN DER OSTEN, CLAUS
; APPLICANT: HALKIER, TORDEN
; APPLICANT: ANDERSEN, CARSTEN
; APPLICANT: BAUDITZ, PETER
; APPLICANT: HANSEN, PETER KAMP
; TITLE OF INVENTION: PROTEASE VARIANTS AND COMPOSITIONS
; FILE REFERENCE: 4946,200-US
; CURRENT APPLICATION NUMBER: US/08/963,851
; CURRENT FILING DATE: 1997-11-04
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
```

```
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
; US-08-963-851-14

Query Match          61.1%; Score 33; DB 4; Length 59;
Best Local Similarity 45.5%; Pred. No. 9.6;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1 BEVFXGMDYS 11
Db      38 EKHIPGLEYS 48
```

Search completed: December 22, 2003, 16:43:41
JOD time : 10.133 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:40:14 ; Search time 20.8667 Seconds
(without alignments)
98.451 Million cell updates/sec

98.451 Million cell updates/sec

Title: US-09-909-164-8

Perfect score: 54

Sequence: 1 EEVVPXGMDYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	66.7	126	US-10-393-840-946	Sequence 946, App
2	36	66.7	937	US-10-289-757-73	Sequence 168, App
3	36	66.7	938	US-10-289-757-168	Sequence 10, App
4	36	66.7	1049	US-10-217-700-10	Sequence 71, App
5	36	66.7	1062	US-10-289-757-71	Sequence 8, App
6	36	66.7	1068	US-10-217-700-8	Sequence 11, App
7	36	66.7	1081	US-10-217-700-4	Sequence 9, App
8	36	66.7	1083	US-10-217-700-11	Sequence 2, App
9	36	66.7	1084	US-10-217-700-9	Sequence 160, App
10	35	64.8	440	US-09-813-408-27	Sequence 58, App
11	34.5	63.9	588	US-10-094-749-1660	Sequence 511, App
12	34	63.0	222	US-10-091-007-58	Sequence 4, App
13	34	63.0	1062	US-09-815-242-5111	Sequence 4, App
14	34	63.0	3472	US-10-029-120-4	Sequence 4, App
15	34	63.0	3472	US-10-027-806-4	Sequence 4, App

16	34	63.0	3472	US-10-034-623-4	Sequence 4, App
17	34	63.0	3472	US-10-027-801-4	Sequence 4, App
18	33	61.1	59	US-09-948-080-14	Sequence 14, App
19	33	61.1	236	US-10-029-386-32076	Sequence 32076, A
20	33	61.1	283	US-09-738-626-4881	Sequence 4881, A
21	33	61.1	299	US-09-815-242-10697	Sequence 10697, A
22	33	61.1	322	US-10-156-761-12418	Sequence 12418, A
23	33	61.1	736	US-09-978-295A-526	Sequence 526, App
24	33	61.1	736	US-09-978-697-526	Sequence 526, App
25	33	61.1	736	US-09-978-192A-526	Sequence 526, App
26	33	61.1	736	US-09-999-832A-526	Sequence 526, App
27	33	61.1	736	US-09-978-189-526	Sequence 526, App
28	33	61.1	736	US-09-978-833A-526	Sequence 526, App
29	33	61.1	736	US-09-978-585A-526	Sequence 526, App
30	33	61.1	736	US-09-978-191A-526	Sequence 526, App
31	33	61.1	736	US-09-978-403A-526	Sequence 526, App
32	33	61.1	736	US-09-978-564A-526	Sequence 526, App
33	33	61.1	736	US-09-999-833A-526	Sequence 526, App
34	33	61.1	736	US-09-981-915A-526	Sequence 526, App
35	33	61.1	736	US-09-978-824-526	Sequence 526, App
36	33	61.1	736	US-09-918-585A-526	Sequence 526, App
37	33	61.1	736	US-09-978-423A-526	Sequence 526, App
38	33	61.1	736	US-09-978-193A-526	Sequence 526, App
39	33	61.1	736	US-09-999-830A-526	Sequence 526, App
40	33	61.1	736	US-09-978-187A-526	Sequence 526, App
41	33	61.1	736	US-09-978-643A-526	Sequence 526, App
42	33	61.1	736	US-09-978-375A-526	Sequence 526, App
43	33	61.1	736	US-09-978-188A-526	Sequence 526, App
44	33	61.1	736	US-09-978-188A-526	Sequence 526, App
45	33	61.1	736	US-09-978-298A-526	Sequence 526, App

ALIGNMENTS

RESULT 1
US-10-393-840-946
Sequence 946, Application US/10393840
Publication No. US2003022922A1
GENERAL INFORMATION:
APPLICANT: Blockberg, Leonard N.
TITLE OF INVENTION: Materials and Methods for the
FILE REFERENCE: 11000.101263
CURRENT APPLICATION NUMBER: US/10/393,840
PRIOR FILING DATE: 2003-03-20
PRIOR APPLICATION NUMBER: US 09/636,800
PRIOR FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: US 09/170,862
PRIOR FILING DATE: 1998-10-13
PRIOR APPLICATION NUMBER: US 60/148,426
PRIOR FILING DATE: 1999-08-11
PRIOR APPLICATION NUMBER: PCT NZ/99/00169
NUMBER OF SEQ ID NOS: 956
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 946
LENGTH: 126
TYPE: PRT
ORGANISM: Pinus radiata
US-10-393-840-946
Query Match
Best Local Similarity 66.7%; Score 36; DB 12; Length 126;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
3 VVPPXGMDYS 11
|||
23 VVPPXGMDYS 31
Db
RESULT 2
US-10-289-757-73


```

; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/10/217,700
; CURRENT FILING DATE: 2002-08-12
; EARLIER APPLICATION NUMBER: 09/394,272
; EARLIER FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1068
; TYPE: PRT
; ORGANISM: Zea mays
US-10-217-700-8

Query Match      66.7%; Score 36; DB 15; Length 1068;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPGXGMDYS 11
Db 435 VVPGXGMDYS 443

RESULT 7
US-10-217-700-4
; Sequence 4, Application US/10217700
; Publication No. US20030070191A1
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/10/217,700
; CURRENT FILING DATE: 2002-08-12
; EARLIER APPLICATION NUMBER: 09/394,272
; EARLIER FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1081
; TYPE: PRT
; ORGANISM: Craterostigma Plantagineum
US-10-217-700-4

Query Match      66.7%; Score 36; DB 15; Length 1081;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPGXGMDYS 11
Db 445 VVPGXGMDYS 453

RESULT 8
US-10-217-700-11
; Sequence 11, Application US/10217700
; Publication No. US20030070191A1
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/10/217,700
; CURRENT FILING DATE: 2002-08-12
; EARLIER APPLICATION NUMBER: 09/394,272
; EARLIER FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11

```

```

; LENGTH: 1083
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-217-700-11

Query Match      66.7%; Score 36; DB 15; Length 1083;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPGXGMDYS 11
Db 483 VVPGXGMDYS 491

RESULT 9
US-10-217-700-9
; Sequence 9, Application US/10217700
; Publication No. US20030070191A1
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/10/217,700
; CURRENT FILING DATE: 2002-08-12
; EARLIER APPLICATION NUMBER: 09/394,272
; EARLIER FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1084
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-217-700-9

Query Match      66.7%; Score 36; DB 15; Length 1084;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPGXGMDYS 11
Db 453 VVPGXGMDYS 461

RESULT 10
US-09-813-408-27
; Sequence 27, Application US/09813408
; Publication No. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Maris, Barry
; TITLE OF INVENTION: Methods For The Synthesis of Polynucleotides And Combinatorial
; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813,408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 27
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Aeropyrum pernix
US-09-813-408-27

Query Match      64.8%; Score 35; DB 11; Length 440;
Best Local Similarity 66.7%; Pred. No. 92;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EVVPGXGMDY 10
Db 120 EVVPGXGMDY 128

```

```
RESULT 11
US-10-094-749-1660
; Sequence 1660, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1660
; LENGTH: 588
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-1660

Query Match      63.9%; Score 34.5; DB 12; Length 588;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Qy      1 EEVVPXGMDY 10
Db      127 EEVVP-GWDF 135

RESULT 12
US-10-091-007-58
; Sequence 58, Application US/10091007
; Publication No. US20030170782A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Le Page, Richard W F      Wells, Jeremy M
; APPLICANT: Hamniffy, Sean B
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PMC/P21978MO
; CURRENT APPLICATION NUMBER: US/10/091,007
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: GB 9921125.2
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 58
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-10-091-007-58

Query Match      63.0%; Score 34; DB 12; Length 222;
Best Local Similarity 50.0%; Pred. No. 68;
```

```
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy      1 EEVVPXGMDY 10
Db      201 KKIVPIGFDY 210

RESULT 13
US-09-815-242-5111
; Sequence 5111, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyckind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5111
; LENGTH: 1062
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5111

Query Match      63.0%; Score 34; DB 9; Length 1062;
Best Local Similarity 65.7%; Pred. No. 3.9e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      5 PXXGMDYS 11
Db      321 PXXGMDYS 327

RESULT 14
US-10-029-120-4
; Sequence 4, Application US/10029120
; Publication No. US20030175708A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCCP.002A
; CURRENT APPLICATION NUMBER: US/10/029,120
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
```

; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-029-120-4

Query Match 63.0%; Score 34; DB 12; Length 3472;
Best Local Similarity 45.5%; Pred. No. 1.5e+03;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
|:|:|:|:|:|:|
Db 2294 EDVIPRGISFS 2304

RESULT 15

US-10-027-806-4
; Sequence 4, Application US/10027806
; Publication No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schlegel, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCRP.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-806-4

Query Match 63.0%; Score 34; DB 14; Length 3472;
Best Local Similarity 45.5%; Pred. No. 1.5e+03;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
|:|:|:|:|:|:|
Db 2294 EDVIPRGISFS 2304

Search completed: December 22, 2003, 17:32:42
Job time : 21.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:24:36 ; Search time 9.06667 Seconds

(without alignments)
116.675 Million cell updates/sec

Title: US-09-909-164-8

Perfect score: 54

Sequence: 1 EEWVPPXGMDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	74.1	156	2	SS4619
2	38	70.4	363	2	D69551
3	36	66.7	102	2	A42452
4	36	66.7	341	2	S72649
5	36	66.7	348	2	S72650
6	36	66.7	460	2	G96764
7	36	66.7	1049	2	UC4783
8	36	66.7	1068	1	JQ1329
9	36	66.7	1081	2	T09837
10	36	66.7	1083	2	T04062
11	36	66.7	1084	2	T04103
12	35	64.8	425	2	T24111
13	35	64.8	433	2	H87660
14	35	64.8	440	2	H72784
15	35	64.8	1150	2	T20173
16	35	64.8	1474	2	F69009
17	35	64.8	2747	2	B49132
18	34	63.0	99	2	S00210
19	34	63.0	155	2	S38255
20	34	63.0	168	2	S58208
21	34	63.0	290	2	AG3104
22	34	63.0	290	2	D98182
23	34	63.0	296	2	D72745
24	34	63.0	357	1	G69280
25	34	63.0	366	2	G69350
26	34	63.0	565	2	E86655
27	34	63.0	587	2	A70164
28	34	63.0	587	2	F81138
29	34	63.0	906	2	T48898

30	34	63.0	908	2	T48899	disease resistance
31	34	63.0	1062	2	F83335	RND multidrug effl
32	34	63.0	1062	2	T30830	hypothetical prote
33	34	63.0	3472	2	T31308	hypothetical 367K
34	33	61.1	97	2	A99427	partial transposas
35	33	61.1	128	2	A90471	hypothetical prote
36	33	61.1	172	2	S27021	fibroblast growth
37	33	61.1	184	2	E90335	hypothetical prote
38	33	61.1	225	2	S57810	hypothetical prote
39	33	61.1	247	2	A96001	conserved hypotnet
40	33	61.1	257	2	A96546	unknown protein (i
41	33	61.1	262	2	F90298	transposase ISCI05
42	33	61.1	267	2	C90307	transposase ISCI05
43	33	61.1	276	2	C64417	hypothetical prote
44	33	61.1	283	2	G83055	pancitate-betaalan
45	33	61.1	299	2	E90487	transposase ISCI05

ALIGNMENTS

RESULT 1

SS4619
hypothetical protein YOR013w - yeast (Saccharomyces cerevisiae)

N/Alternate names: hypothetical protein O2612; hypothetical protein YOR0303.3

C/Species: Saccharomyces cerevisiae

C/Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002

C/Accession: SS4619; S66879

R/de Haan, M.; Maarse, A.C.; Grivell, L.A.
submitted to the EMBL Data Library, May 1995

A/Reference number: SS4617

A/Accession: SS4619

A/Molecule type: DNA

A/Residues: 1-156 <DEH>

A/Cross-references: EMBL:X87331; NID:G1041652; PIDN:CAA60762.1; PID:G829123

R/de Haan, M.; Grivell, L.A.; Maarse, A.C.
submitted to the Protein Sequence Database, July 1996

A/Reference number: S66877

A/Accession: S66879

A/Molecule type: DNA

A/Residues: 1-156 <DEW>

A/Cross-references: EMBL:Z74920; NID:G1420109; PIDN:CAA9201.1; PID:G1420111; MIPS:YOR

C/Genetics:

A/Cross-references: SGD:S0005539

A/Map position: 15R

C/Superfamily: hypothetical protein YOR013w

Query Match 74.1% Score 40; DB 2; Length 156;
Best Similarity 77.8% Pred. No. 1;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EEWVPPXGMDY 10

DB 50 EEWVPPXGMDY 58

RESULT 2

D69551
conserved hypothetical protein AP2411 - Archaeoglobus fulgidus

C/Species: Archaeoglobus fulgidus

C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999

C/Accession: D69551

R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dods

; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kinkness, E.

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A/Authors: Uitterlinden, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes,

A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archa

A/Reference number: A69250; WUID:98049343; PMID:9389475

A/Accession: D69551

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA
 A: Residues: 1-363 <XLE>
 A: Cross-references: GB:AE001109; GB:AE000782; NID:g2689432; PIDN:BA091255.1; PID:g265066

Query Match 70.4%; Score 38; DB 2; Length 363;
 Best Local Similarity 54.5%; Pred. No. 6.8;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 1 BEVVPXGMDYS 11
 |||
 Db 120 ENIVPGIDFS 130

RESULT 3

A42452
 VI protein - tobacco yellow dwarf virus (strain Australia)

C: Species: tobacco yellow dwarf virus
 C: Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
 C: Accession: A42452

R: Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.

A: Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow

A: Reference number: A42452; MUID:92188538; PMID:1546458

A: Accession: A42452

A: Molecule type: DNA

A: Residues: 1-102 <MOR>

A: Cross-references: GB:M81103; NID:g335283; PIDN:AAA47947.1; PID:g335284

Query Match 66.7%; Score 36; DB 2; Length 102;
 Best Local Similarity 60.0%; Pred. No. 4.3;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Oy 2 EVVVPXGMDYS 11
 |||
 Db 7 QVPSGINTS 16

RESULT 4

S72649
 sucrose-phosphate synthase (EC 2.4.1.14) isoform 2 - Citrus unshiu (fragment)

C: Species: Citrus unshiu

C: Date: 24-Oct-1998 #sequence_revision 24-Oct-1998 #text_change 21-Jul-2000

C: Accession: S72649

R: Komatsu, A.; Takano, Y.; Omura, M.; Akihama, T.

Mol. Gen. Genet. 252, 346-351, 1996

A: Title: Cloning and molecular analysis of cDNAs encoding three sucrose phosphate synthase

A: Reference number: S72648; MUID:96439842; PMID:8842155

A: Accession: S72649

A: Status: nucleic acid sequence not shown

A: Molecule type: mRNA

A: Residues: 1-341 <KOM>

A: Cross-references: EMBL:AB006319; NID:g2588891; PIDN:BA023215.1; PID:g2588892

A: Experimental source: fruit, cv. Miyagawa-Wase

A: Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1997

C: Genetics:

A: Gene: SP52

C: Function:

A: Description: catalyzes formation of sucrose-6-phosphate from UDPglucose and D-fructose

A: Pathway: sucrose biosynthesis

C: Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology

C: Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis

F: 1-341/Domain: sucrose/sucrose-phosphate synthase homology (fragment) <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 341;
 Best Local Similarity 66.7%; Pred. No. 16;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 3 VVPXGMDYS 11
 |||
 Db 228 VIPPGMDFS 236

RESULT 5

S72650

sucrose-phosphate synthase (EC 2.4.1.14) isoform 3 - Citrus unshiu (fragment)

C: Species: Citrus unshiu

C: Date: 24-Oct-1998 #sequence_revision 24-Oct-1998 #text_change 21-Jul-2000

C: Accession: S72650

R: Komatsu, A.; Takano, Y.; Omura, M.; Akihama, T.

Mol. Gen. Genet. 252, 346-351, 1996

A: Title: Cloning and molecular analysis of cDNAs encoding three sucrose phosphate synt

A: Reference number: S72648; MUID:96439842; PMID:8842155

A: Accession: S72650

A: Status: nucleic acid sequence not shown

A: Molecule type: mRNA

A: Residues: 1-348 <KOM>

A: Cross-references: EMBL:AB006660; NID:g2351059; PIDN:BA022071.1; PID:g2351060

A: Experimental source: fruit, cv. Miyagawa-Wase

A: Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1997

C: Genetics:

A: Gene: SP53

C: Function:

A: Description: catalyzes formation of sucrose-6-phosphate from UDPglucose and D-fructo

C: Superfamily: sucrose biosynthesis

C: Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis

F: 1-348/Domain: sucrose/sucrose-phosphate synthase homology (fragment) <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 348;
 Best Local Similarity 66.7%; Pred. No. 17;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 3 VVPXGMDYS 11
 |||
 Db 234 VIPPGMDFS 242

RESULT 6

G96764

unknown protein F25P22.17 [imported] - Arabidopsis thaliana

C: Species: Arabidopsis thaliana (mouse-ear cress)

C: Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C: Accession: G96764

R: Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alons

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K

ansen, N.F.; Hughes, B.; Hultzer, L.

Nature 408, 816-820, 2000

A: Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Marti, R.; Marziani

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A: Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A: Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A: Reference number: A86141; MUID:21016719; PMID:11130712

A: Accession: G96764

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-460 <STO>

A: Cross-references: GB:AE005173; NID:g6692750; PIDN:AAF24856.1; GSPDB:GN00141

C: Genetics:

A: Gene: F25P22.17

A: Map position: 1

Query Match 66.7%; Score 36; DB 2; Length 460;
 Best Local Similarity 70.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 BEVVPXGMDY 10
 |||
 Db 218 BEVPSAMDY 227

RESULT 7

JC4783

sucrose-phosphate synthase (EC 2.4.1.14) - rice

C: Species: Oryza sativa (rice)

C>Date: 10-May-1996 #sequence_revision 16-Aug-1996 #text_change 18-Jun-1999
 C/Accession: J04783
 R.Valdez-Alarcon, U.J.; Ferrando, M.; Saleno, G.; Jimenez-Morilla, B.; Herrera-Betrelle
 Gene 170, 217-222, 1996
 A>Title: Characterization of a rice sucrose-phosphate synthase-encoding gene.
 A/Reference number: J04783; MUID:96253138; PMID:8666248
 A/Accession: J04783
 A/Molecule type: mRNA
 A/Residues: 1-1049 <VAL>
 A/Cross-references: GB:U31175; NID:91449931; PIDN:AA049379.1; PID:9988270
 A/Note: UDPglucose-fructose-phosphate glucosyltransferase; Sucrosephosphate-UDPglucose
 C/Comment: This enzyme catalyzes the formation of sucrose-phosphate form UDP-glucose and
 C/Genetics:
 A/Gene: Spst
 A/Introns: 24/1, 103/3, 183/3, 205/3, 435/3, 475/3, 519/3, 578/3, 596/3, 617/3, 931/3, 9
 C/Function:
 A>Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fruc
 A/Pathway: sucrose biosynthesis
 C/Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
 C/Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis
 P:180-663/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1049;
 Best Local Similarity 66.7%; Pred. No. 58;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 VWPXGMDYS 11
 DB 436 VWPXGMDYS 444

RESULT 8
 J01329
 sucrose-phosphate synthase (EC 2.4.1.14) - maize
 C/Species: Zea mays (maize)
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C/Accession: J01329; PQ0260
 R.Morrell, A.C.; Brunau, J.M.; Summerfelt, K.; Boerig, M.; Voelker, T.A.
 Plant Cell 3, 1121-1130, 1991
 A>Title: Expression of a maize sucrose phosphate synthase in tomato alters leaf carbohyd
 A/Reference number: J01329; MUID:92338837; PMID:1840396
 A/Accession: J01329
 A/Molecule type: mRNA
 A/Residues: 1-1068 <MOR>
 A/Cross-references: GB:M97550; NID:9166625; PIDN:AAA3513.1; PID:9166626
 A/Accession: PQ0260
 A/Molecule type: protein
 A/Residues: 71-74;206-212;471-481;872-892 <MOR1>
 C/Comment: This enzyme transfers the glucosyl group from UDPglucose to fructose-6-phosph
 C/Comment: This enzyme is involved in the regulation of carbon partitioning in the leave
 C/Function:
 A>Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fruc
 A/Pathway: sucrose biosynthesis
 C/Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
 C/Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis
 P:178-666/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 1; Length 1068;
 Best Local Similarity 66.7%; Pred. No. 59;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 VWPXGMDYS 11
 DB 435 VWPXGMDYS 443

RESULT 9
 T09837
 sucrose-phosphate synthase (EC 2.4.1.14) isoform 2 - Craterostigma plantaginum
 C/Species: Craterostigma plantaginum
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C/Accession: T09837
 R.Ingram, J.; Chandler, J.W.; Gallagher, L.; Salamini, F.; Bartels, D.

Plant Physiol. 115, 113-121, 1997
 A>Title: Analysis of cDNA clones encoding sucrose-phosphate synthase in relation to s
 A/Reference number: Z16874; MUID:97451773; PMID:9306694
 A/Accession: T09837
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-1081 <ING>
 A/Cross-references: EMBL:Y11795; NID:92190349; PIDN:CAA72491.1; PID:92190350
 A/Experimental source: ABA-treated callus
 C/Genetics:
 A/Gene: spst
 C/Function:
 A>Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-f
 A/Pathway: sucrose biosynthesis
 C/Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
 C/Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis
 P:176-674/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1081;
 Best Local Similarity 66.7%; Pred. No. 60;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 VWPXGMDYS 11
 DB 445 VWPXGMDYS 453

RESULT 10
 T04062
 sucrose-phosphate synthase homolog F28M11.40 - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 16-Jul-1999
 C/Accession: T04062
 R.Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.I
 submitted to the Protein Sequence Database, March 1999
 A/Reference number: Z15184
 A/Accession: T04062
 A/Molecule type: DNA
 A/Residues: 1-1083 <BEV>
 A/Cross-references: EMBL:AL049487
 A/Experimental source: cultivar Columbia; BAC clone F28M11
 C/Genetics:
 A/Map position: 4
 A/Introns: 86/3, 116/3, 255/3, 322/2, 482/3, 526/3, 570/3, 629/3, 647/3, 668/3, 949/3,
 A/Note: F28M11.40
 C/Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
 P:230-714/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1083;
 Best Local Similarity 66.7%; Pred. No. 60;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 VWPXGMDYS 11
 DB 483 VWPXGMDYS 491

RESULT 11
 T04103
 sucrose-phosphate synthase (EC 2.4.1.14) 1 - rice
 C/Species: Oryza sativa (rice)
 C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
 C/Accession: T04103
 R.Sakamoto, M.; Satozawa, T.; Kishimoto, N.; Higo, K.; Shimada, H.; Fujimura, T.
 Plant Sci. 112, 207-217, 1995
 A>Title: Structure and RFLP mapping of a rice sucrose phosphate synthase (SPS) gene cl
 A/Reference number: Z15212
 A/Accession: T04103
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-1084 <SAK>
 A/Cross-references: EMBL:D45890; PIDN:BA08304.1
 A/Experimental source: subsp. Japonica

C:Genetics:
 A:Gene: Spst1
 A:Map position: 1
 A:introns: 120/3; 200/2; 221/3; 452/3; 492/3; 536/3; 595/3; 613/3; 634/3; 946/3; 989/2
 C:Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
 C:Keywords: glycoyltransferase; hexosyltransferase
 F:196-680/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1084;
 Best Local Similarity 66.7%; Pred. No. 60;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 3 VVPXGMDYS 11
 ||:|||||
 Db 453 VVPXGMDYS 461

RESULT 12

T24111
 hypothetical protein R10D12.10 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T24111
 R:Percy, C.

submitted to the EMBL Data Library, October 1996
 A:Reference number: Z19842

A:Accession: T24111
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-425 <WIL>
 A:Cross-references: EMBL:Z81109; PIDN:CAB03241.1; GSPDB:GN00023; CBSP:R10D12.10
 A:Experimental source: clone R10D12
 C:Genetics:
 A:Gene: CBSP:R10D12.10
 A:Map position: 5
 A:introns: 23/3; 56/3; 113/3; 257/2

Query Match 64.8%; Score 35; DB 2; Length 425;
 Best Local Similarity 50.0%; Pred. No. 34;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 1 EEVVPXGMDY 10
 ||:|||||
 Db 335 EQIVPGGLQY 344

RESULT 13

H87660
 peptidoglycan-binding protein, probable [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: H87660

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Iatb, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 N.; J.; Esmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001.

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: H87660

A:Molecule type: DNA

A:Status: preliminary

A:Residues: 1-433 <STO>

A:Cross-references: GB:A8005673; NID:g13425020; PIDN:AAK25284.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC3322

Query Match 64.8%; Score 35; DB 2; Length 433;
 Best Local Similarity 54.5%; Pred. No. 34;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Oy 1 EEVVPXGMDYS 11
 ||:|||||
 Db 266 EVILPPGFDS 276

RESULT 14

H72784

probable alkaline proteinase APE0263 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C:Accession: H72784

R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Tak
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.;

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero

A:Reference number: A72450; MUID:9310339; PMID:10382966

A:Accession: H72784

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-440 <KAM>

A:Cross-references: DDBJ:AP000058; NID:g5103388; PIDN:BA479178.1; PID:g5103657

A:Experimental source: strain K1

C:Genetics:

A:Gene: APE0263

C:Superfamily: subtilisin; subtilisin homology

Query Match 64.8%; Score 35; DB 2; Length 440;
 Best Local Similarity 66.7%; Pred. No. 35;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 2 EVVPXGMDY 10
 ||:|||||
 Db 120 EVLPWGVYD 128

RESULT 15

T20173

hypothetical protein CS3A5.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C:Accession: T20173; T23857

R:Mortimore, B.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19232

A:Accession: T20173

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1150 <WIL>

A:Cross-references: EMBL:Z81486; PIDN:CAB03994.1; GSPDB:GN00023; CBSP:CS3A5.2

A:Experimental source: clone CS3A5

R:Matthews, L.

submitted to the EMBL Data Library, August 1996

A:Reference number: Z19808

A:Accession: T23857

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1150 <WIL>

A:Cross-references: EMBL:Z78015; PIDN:CAB01437.1; GSPDB:GN00023; CBSP:CS3A5.2

A:Experimental source: clone R02D5

C:Genetics:

A:Gene: CBSP:CS3A5.2

A:introns: 33/3; 63/3; 132/3; 169/3; 221/3; 299/3; 379/2; 423/2; 438/2; 471/1; 513/2;

Query Match 64.8%; Score 35; DB 2; Length 1150;
 Best Local Similarity 66.7%; Pred. No. 1e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 3 VVPXGMDYS 11
 ||:|||||
 Db 562 VLPVGIIDYS 570

Search completed: December 22, 2003, 17:44:57
 Job time : 11.0667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:43:51 / Search time 4.6 Seconds

(without alignments)
112.455 Million cell updates/sec

Title: US-09-909-164-8
Perfect score: 54
Sequence: 1 EEVVPXGMDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing filter 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	70.4	1058	1	CARB_FUSNN
2	36	66.7	102	1	YILK_TYDVA
3	36	66.7	1049	1	SPS_ORYSA
4	36	66.7	1068	1	SPS_MAIZE
5	36	66.7	1081	1	SPS2_CRAPL
6	35	66.8	2778	1	FAR_DROME
7	34.5	62.9	748	1	KHL1_HUMAN
8	34	63.0	154	1	PLAS_ORYSA
9	34	63.0	155	1	PLAS_HORVU
10	34	63.0	168	1	PLAT_POPNI
11	34	63.0	566	1	SYFB_BORBU
12	34	63.0	908	1	RBL4_ARATH
13	34	63.0	908	1	RPP8_ARATH
14	34	63.0	910	1	RPP8_ARATH
15	33	61.1	276	1	Y939_METUA
16	33	61.1	283	1	PANC_PSEAE
17	33	61.1	394	1	HMPA_VIBCH
18	33	61.1	421	1	ACDM_RAT
19	33	61.1	423	1	ECB2_HALEL
20	33	61.1	423	1	ECB1_HALEL
21	33	61.1	787	1	ECR2_HUMAN
22	33	61.1	801	1	FCR3_MOUSE
23	33	61.1	806	1	CEK2_CHICK
24	33	61.1	877	1	SUDH_SCHPO
25	33	61.1	982	1	ENV_SFV3L
26	33	61.1	1401	1	RPOC_VIBCH
27	33	61.1	2717	1	ZEP1_HUMAN
28	32.5	60.2	472	1	ET2A_XENLA
29	32	59.3	97	1	PLAS_DAUCA
30	32	59.3	165	1	TPX_LISIN
31	32	59.3	165	1	TPX_LISMO
32	32	59.3	175	1	HES3_RAT
33	32	59.3	231	1	ARAD_ECOLI

34	32	59.3	231	1	ARAD_SALTY	P06190 salmonella
35	32	59.3	233	1	HIS9_THEMA	O9wzr1 thermotoga
36	32	59.3	288	1	CGD2_RAT	O04827 rattus norv
37	32	59.3	289	1	CGD2_HUMAN	P30279 homo sapien
38	32	59.3	289	1	CGD2_MOUSE	P30280 mus musculu
39	32	59.3	291	1	CGD1_BRARE	O90459 brachydanto
40	32	59.3	291	1	CGD1_XENLA	P50755 xenopus lae
41	32	59.3	291	1	CGD2_CHICK	P49706 gallus gall
42	32	59.3	291	1	CGD2_XENLA	P53782 xenopus lae
43	32	59.3	292	1	CGD1_CHICK	P55169 gallus gall
44	32	59.3	292	1	CGD3_HUMAN	P30281 homo sapien
45	32	59.3	295	1	CGD1_HUMAN	P24385 homo sapien

ALIGNMENTS

RESULT 1
ID CARB_FUSNN STANDARD; PRT; 1058 AA.
AC O8RG86;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cardamoyl-phosphate synthase large chain (EC 6.3.5.5) (Cardamoyl-phosphate synthetase ammonia chain).
GN CARB_OR_FND422.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteriia; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=2186394; PubMed=11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A., Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L., Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman R., Bernal A., Laren N., D'Souza M., Malinas T., Pusch G., Haseikorn R., Fomstein M., Kyridis N., Overbeek R.;
RA "Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586.";
RA J. Bacteriol. 184:2005-2018 (2002).
CC -1- CARBATIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + cardamoyl phosphate.
CC -1- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -1- PATHWAY: Arginine biosynthesis.
CC -1- PATHWAY: Pyrimidine biosynthesis; first step.
CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize cardamoyl phosphate (By similarity).
CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AE010554; AAL94625.1; ALT_INIT.
CC HAMAP; MF_01210; -; 1.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005483; CPhase_L.
DR InterPro; IPR005479; CPhase_L_D2.
DR InterPro; IPR005480; CPhase_L_D3.
DR InterPro; IPR005481; CPhase_L_N.
DR InterPro; IPR004362; MGS_1ike.
DR Pfam; PF00289; CPhase_L_D2; 2.
DR Pfam; PF02786; CPhase_L_D2; 2.
DR Pfam; PF02787; CPhase_L_D3; 1.

PFam: PF02142; MCS: 1.
 DR PRINTS: PR00098; CPSASR.
 DR TIGRPFAM; TIGR01369; CPSASE1_1tr; 1.
 DR PROSITE; PS00866; CPSASE_1; 2.
 DR PROSITE; PS00867; CPSASE_2; 2.
 KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 KW ATP-binding; Manganese; Complete proteome.
 FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
 FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
 FT REPEAT 930 1058 ALLOSTERIC DOMAIN.
 FT REPEAT 1 546
 FT REPEAT 547 1058
 FT NP_BIND 153 210 ATP (POTENTIAL).
 FT NP_BIND 302 352 ATP (POTENTIAL).
 FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
 FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
 FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
 FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
 FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
 SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1E39F CRC64;

Query Match 70.4%; Score 38; DB 1; Length 1058;
 Best Local Similarity 60.0%; Pred. No. 9.1;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 2 EVVPGMDYS 11
 Db 190 EIVPGMLNYS 199

RESULT 2
 Y11K_TYDVA STANDARD; PRT; 102 AA.

AC P31619;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-1993 (Rel. 27, Last annotation update)
 DE Hypothetical 11.2 kDa protein.
 GN V1.
 OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
 OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
 CX NCBI_TaxID=31599;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92188538; PubMed=1546458;
 RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;
 RT "The nucleotide sequence of the infectious cloned DNA component of
 tobacco yellow dwarf virus reveals features of geminiviruses
 infecting monocotyledonous plants.";
 RT infecting monocotyledonous plants.";
 RL Virology 187:633-642(1992).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M81103; AAA47947.1; -
 DR PIR; A42452; A42452.
 DR InterPro; IPR002621; Gemini1_mov.
 DR Pfam; PF01708; Gemini1_mov; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF55B67 CRC64;

Query Match 66.7%; Score 36; DB 1; Length 102;
 Best Local Similarity 60.0%; Pred. No. 2;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 2 EVVPGMDYS 11
 Db 190 EIVPGMLNYS 199

Db 7 QVPSGINYS 16

RESULT 3
 SPS_ORYSA STANDARD; PRT; 1049 AA.
 ID SPS_ORYSA
 AC Q43802;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Sucrose-phosphate synthase (EC 2.4.1.14)
 DE (UDP-glucose-fructose-phosphate glucosyltransferase).
 DE Oryza sativa (Rice).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriactoidae; Oryzaceae; Oryza.
 CX NCBI_TaxID=4530;
 RP SEQUENCE FROM N.A.
 RX STRAIN=cv. Indica-IR36; Tissue=leaf;
 RX MEDLINE=96235138; PubMed=8666248;
 RA Valdez-Alarcon J.J., Ferrando M., Jimenez-Morales B.,
 RA Herrera-Estrella L.;
 RT "Characterization of a rice sucrose-phosphate synthase-encoding
 RT gene.";
 RL Gene 170:217-222(1996).
 CC -----
 CC -1- FUNCTION: INVOLVED IN THE REGULATION OF CARBON PARTITIONING IN
 CC THE LEAVES OF PLANTS. MAY REGULATE THE SYNTHESIS OF SUCROSE AND
 CC THEREFORE PLAY A MAJOR ROLE AS A LIMITING FACTOR IN THE EXPORT OF
 CC PHOTOSIMILATES OUT OF THE LEAF.
 CC -1- CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP +
 CC sucrose 6-phosphate.
 CC -1- ENZYME REGULATION: ACTIVITY REGULATED BY PHOSPHORYLATION AND
 CC MODERATED BY CONCENTRATION OF METABOLITES AND LIGHT.
 CC -1- PATHWAY: Sucrose synthesis.
 CC -1- SUBUNIT: Homodimer or homotrimer (By similarity).
 CC -1- PTM: PHOSPHORYLATED. HOWEVER, PHOSPHORYLATION IS NOT ESSENTIAL FOR
 CC ENZYME FUNCTION (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U33175; AAC49379.1; -
 DR PIR; JC4783; JC4783.
 DR Gramene; Q43802; -
 DR InterPro; IPR001296; Glyco_transf_1.
 DR Pfam; PF00534; Glycosyltransf_1; 1.
 DR Transferrase; Glycosyltransferase; Phosphorylation.
 FT DOMAIN 22 29 POLY-GLY.
 FT DOMAIN 695 698 POLY-GLY.
 FT DOMAIN 775 779 POLY-ARG.
 SQ SEQUENCE 1049 AA; 116455 MW; ED862E2819AA4B04 CRC64;

Query Match 66.7%; Score 36; DB 1; Length 1049;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 VVPGMDYS 11
 Db 436 VVPGMDYS 444

RESULT 4
 SPS_MAIZE STANDARD; PRT; 1068 AA.
 ID SPS_MAIZE
 AC P31927;
 DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE Sucrose-phosphatase synthase (EC 2.4.1.14) (UDP-glucose-fructose-
 phosphate glucosyltransferase).
 GN SPS.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
 NCBI_TaxID=4577;
 RX MEDLINE=92238837; PubMed=1840396;
 RA Morrell A.C., Brunau J.-M., Summerfelt K., Boersig M., Voelker T.A.;
 RT "Expression of a maize sucrose phosphate synthase in tomato alters
 leaf carbohydrate partitioning."
 RL Plant Cell 3:1121-1130(1991).
 CC -1- FUNCTION: INVOLVED IN THE REGULATION OF CARBON PARTITIONING IN
 THE LEAVES OF PLANTS. MAY REGULATE THE SYNTHESIS OF SUCROSE AND
 THEREFORE PLAY A MAJOR ROLE AS A LIMITING FACTOR IN THE EXPORT OF
 PHOTOSYNTHETATES OUT OF THE LEAF.
 CC -1- CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP +
 sucrose 6-phosphate.
 CC -1- ENZYME REGULATION: ACTIVITY REGULATED BY PHOSPHORYLATION AND
 MODERATED BY CONCENTRATION OF METABOLITES AND LIGHT.
 CC -1- PATHWAY: Sucrose synthesis.
 CC -1- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
 CC -1- DEVELOPMENTAL STAGE: GERMINATING SEEDS OR MATURE LEAVES.
 CC -1- PTM: PHOSPHORYLATED. HOWEVER, PHOSPHORYLATION IS NOT ESSENTIAL FOR
 ENZYME FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; M97550; AAA3513.1; -
 DR PIR; J01329; J01329.
 DR MAZEDB; 25294; -
 DR InterPro; IPR001296; Glyco_transf_1.
 DR Pfam; PF00534; Glycosyltransferase; Phosphorylation.
 KM Transferrase; Glycosyltransferase; Phosphorylation.
 FT DOMAIN 25 31 POLY-GLY.
 SQ SEQUENCE 1068 AA; 118575 MW; 074679B5E9A1D282 CRC64;
 Query Match 66.7%; Score 36; DB 1; Length 1068;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 3 VVPXGMDYS 11
 DB 435 VVPGMDPS 443
 RESULT 5
 SPS2_CRAPL STANDARD; PRT; 1081 AA.
 AC 004933;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Sucrose-phosphate synthase 2 (EC 2.4.1.14) (UDP-glucose-fructose-
 phosphate glucosyltransferase 2).
 GN SPS2.
 OS Crataegia plantagineum.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; Lamiales; Lamiales; Lamiales incertae sedis; Toreniceae;
 OC Crataegia.
 OC NCBI_TaxID=4153;
 RX MEDLINE=97451773; PubMed=9306694;
 RA Ingram J., Chandler J.W., Gallagher L., Salami F., Bartels D.;
 RT "Analysis of cDNA clones encoding sucrose-phosphate synthase in
 relation to sugar interconversions associated with dehydration in the
 resurrection plant Crataegia plantagineum Hochst."
 RL Plant Physiol. 115:113-121(1997).
 CC -1- FUNCTION: INVOLVED IN THE REGULATION OF CARBON PARTITIONING IN
 THE LEAVES OF PLANTS. MAY REGULATE THE SYNTHESIS OF SUCROSE AND
 THEREFORE PLAY A MAJOR ROLE AS A LIMITING FACTOR IN THE EXPORT OF
 PHOTOSYNTHETATES OUT OF THE LEAF.
 CC -1- CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP +
 sucrose 6-phosphate.
 CC -1- ENZYME REGULATION: ACTIVITY REGULATED BY PHOSPHORYLATION AND
 MODERATED BY CONCENTRATION OF METABOLITES AND LIGHT.
 CC -1- PATHWAY: Sucrose synthesis.
 CC -1- SUBUNIT: Homodimer or homotetramer (by similarity).
 CC -1- PTM: PHOSPHORYLATED. HOWEVER, PHOSPHORYLATION IS NOT ESSENTIAL FOR
 ENZYME FUNCTION (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; Y11795; CAA72491.1; -
 DR PIR; T09837; T09837.
 DR InterPro; IPR001296; Glyco_transf_1.
 DR Pfam; PF00534; Glycosyltransferase; Phosphorylation.
 KM Transferrase; Glycosyltransferase; Phosphorylation.
 FT DOMAIN 245 248 POLY-SER.
 FT DOMAIN 256 264 POLY-GLU.
 FT DOMAIN 787 790 POLY-ARG.
 SQ SEQUENCE 1081 AA; 120933 MW; DD142DC2F1A72900 CRC64;
 Query Match 66.7%; Score 36; DB 1; Length 1081;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 3 VVPXGMDYS 11
 DB 445 VVPGMDPS 453
 RESULT 6
 FAF_DROME STANDARD; PRT; 2778 AA.
 AC P55824; Q9V9T6; Q9V027;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable ubiquitin carboxyl-terminal hydrolase FAF (EC 3.1.2.15)
 DE (Ubiquitin thiolesterase FAF) (Ubiquitin-specific processing protease
 FAF) (Deubiquitinating enzyme FAF) (Fat facets protein).
 GN FAF OR BCDNA.LD22582 OR CG1945.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RX MEDLINE=93202020; PubMed=1295747;
 RA Fischer-Vize J.A., Rubin G.M., Lehmann R.;

RT "The fa⁺ facets gene is required for Drosophila eye and embryo development.";

RL Development 116:985-1000(1992).

[2]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.R., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutcliffe G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.F., Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D., Man K.H., Doyle C., Baxter E.G., Holt G., Nelson C.R., Miklos G.L.G., Abril J.F., Aghayani A., An H.J., Andrews-Pennkoob C., Baldwin D., Ballew R.M., Baau A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Berns P.V., Bernan B.P., Bhandari D., Bolshakov S., Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P., Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablo B., Delcher A., Deng Z., Mayas A.D., Dew I., Dietz S.M., Doudan K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Flosser C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaser K., Godke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iqbal C., Jatalin M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z., Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Mikhina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclab J.M., Palazzolo M., Peltman G.S., Pan S., Pollard J., Puti V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wassenaar D.A., Weinstein G.M., Weissbach J., Williams S.M., Wodade T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yah R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RA "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

[3]

RP REVISIONS, AND ALTERNATIVE SPLICING.

RC STRAIN=Berkley;

RX MEDLINE=22426069; PubMed=12537572;

RA Madera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S., Hirschsky P., Huang Y., Kaminig J.S., Milburn G.H., Prochuk S.E., Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Bernan B.P., Bettencourt B.R., Celnikier S.E., de Grey A.D.N.J., Drysdale R.A., Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q., Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M., Lewis S.E.;

RA "Annotation of the Drosophila melanogaster euchromatic genome: a systematic review.";

RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).

[4]

RP SEQUENCE OF 1089-2778 FROM N.A. (ISOFORM 1).

RC STRAIN=Berkley;

RX MEDLINE=20196012; PubMed=10731138;

RA Rubin G.M., Hong L., Brokstein P., Evans-Holm M., Frie E., Stapleton M., Harvey D.A.;

RT "A Drosophila complementary DNA resource.";

RL Science 287:2222-2224(2000).

-I- FUNCTION: REQUIRED FOR EYE AND EMBRYO DEVELOPMENT, AND PLAYS A ROLE IN COMPOUND EYE ASSEMBLY AND OCOGENESIS RESPECTIVELY. IN THE LARVAL EYE DISKS, CELLS OUTSIDE THE ASSEMBLING FACETS REQUIRE THE PROTEIN FOR SHORT-RANGE CELL INTERACTIONS THAT PREVENT THE MYSTERY CELLS FROM BECOMING PHOTORECEPTORS. IT IS ALSO REQUIRED FOR NUCLEAR MIGRATION AND CELLULARIZATION IN EARLY EMBRYOGENESIS AND

```

CC CC CANON PLAY A ROLE IN POLE CELL DETERMINATION, DEVELOPMENT OR
CC CC FUNCTION.
CC CC -1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC CC ubiquitin + a thiol.
CC CC -1- ALTERNATIVE PRODUCTS:
CC CC Event-Alternative splicing; Named isoforms=3;
CC CC Comment=Experimental confirmation may be lacking for some
CC CC isoforms;
CC CC Name=1;
CC CC IsoId=P55824-1; Sequence=Displayed;
CC CC Name=2;
CC CC IsoId=P55824-2; Sequence=VSP_005270;
CC CC Name=3;
CC CC IsoId=P55824-3; Sequence=VSP_005269;
CC CC -1- TISSUE SPECIFICITY: EYE DISKS AND OVARIES.
CC CC -1- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
CC CC -1- SIMILARITY: Belongs to peptidase family C19.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb.ch/announce/
CC CC or send an email to license@isb-eb.ch).
CC CC -----
CC CC EMBL; L04959; AAF01345.1; -
CC CC EMBL; L04958; AAF01346.1; -
CC CC EMBL; L04960; AAF01347.1; -
CC CC EMBL; L04960; AAF01348.1; -
CC CC EMBL; AE003779; AAF57198.1; -
CC CC EMBL; AE003779; AANI4291.1; -
CC CC EMBL; AF145677; AAD38652.1; -
CC CC MEROPS; C19.007; -
CC CC DR FLYBASE; FBgn005632; faf.
CC CC DR GO; GO:0005737; C:cytoplasm; IDA.
CC CC DR GO; GO:0007349; P:cellularization; IMP.
CC CC DR GO; GO:0009795; P:embryonic morphogenesis; IMP.
CC CC DR GO; GO:0008583; P:myotery cell fate differentiation (sensu Dr. . .; IMP.
CC CC DR GO; GO:0007097; P:nuclear migration; IMP.
CC CC DR GO; GO:0006512; P:ubiquitin cycle; IGI.
CC CC InterPro; IPR001394; UCH-2.
CC CC DR Pfam; PF00443; UCH_1.
CC CC DR PROSITE; PS00972; UCH_2_1; 1.
CC CC DR PROSITE; PS00973; UCH_2_2; 1.
CC CC DR PROSITE; PS02035; UCH_2_3; 1.
CC CC Ubl conjugation pathway; Hydrolyase; Thiol protease;
CC CC Developmental protein; Vision; Alternative splicing.
CC CC FT ACT_SITE 1677 1677
CC CC FT ACT_SITE 1978 1978
CC CC FT ACT_SITE 1986 1986
CC CC FT ACT_SITE 2705 2778
CC CC FT VARSPLIC 2705 2778
CC CC FT VARSPLIC 2742 2778
CC CC FT IATPATEPMSSELTIVKNSLIISOENPQAKSLQ ->
CC CC FT IATPATEPMSSELTIVKNSLIISOENPQAKSLQ ->
CC CC FT SORCOL (in isoform 2).
CC CC FT /FTID=VSP_005270.
CC CC FT E -> D (IN REF. 1).
CC CC FT T -> S (IN REF. 1; AAF01345).
CC CC FT SEQUENCE 2778 AA; 311139 MW; FFB90438BA53A02B CRC64;
CC CC -----
CC CC Query Match 64.8%; Score 35; DB 1; Length 2778;
CC CC Best Local Similarity 54.5%; Pred.No. 1.1e+02;
CC CC Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0.
CC CC -----
CC CC 1 EEVVPKGMDS 11
CC CC :|||:
CC CC 1394 EVIVPDGQDFS 1404

```

ID KHL1 HUMAN STANDARD; PRT; 748 AA.
 AC Q9NR64; Q9H4X4; Q9NR65; Q9P238;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Kelch-like protein 1.
 GN KHL1 OR KIAA1490.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20347694; PubMed=10886605;
 RA Koob M.D., Nemes J.P., Benzow K.A.;
 RT "The SCAB transcript is an antisense RNA to a brain-specific
 transcript encoding a novel actin-binding protein (KHL1).";
 RL Hum. Mol. Genet. 9:1543-1551(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Brain;
 RC MEDLINE=20277482; PubMed=10819331;
 RA Nagase T., Kikuno R., Ichikawa K.-I., Hirosewa M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XVII.
 The complete sequences of 100 new cDNA clones from brain which code
 for large proteins in vitro.";
 RL DNA Res. 7:143-150(2000).
 RN [3]
 RP SEQUENCE OF 179-409 FROM N.A.
 RA Kay M.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY PLAY A ROLE IN ORGANIZING THE ACTIN CYTOSKELETON OF
 THE BRAIN CELLS.
 CC -1- TISSUE SPECIFICITY: Highly expressed in brain.
 CC -1- SIMILARITY: Contains 1 BTB/POZ domain.
 CC -1- SIMILARITY: Contains 6 Kelch repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF252283; AAF81719.1; -;
 DR EMBL; AF252279; AAF81716.1; -;
 DR EMBL; AB040923; BAA96014.1; ALT_INIT.
 DR EMBL; AL353738; CAC16128.1; -;
 DR Genew; HGNC:6352; KHL1.
 DR MIM; 605332; -;
 DR GO; GO:0003779; F:actin binding activity; NAS.
 DR GO; GO:0030036; P:actin cytoskeleton organization and biogenesis; NAS.
 DR InterPro; IPR000210; BTB_POZ.
 DR InterPro; IPR006552; Kelch_rep.
 DR Pfam; PF00651; BTB; 1.
 DR Pfam; PF03344; Kelch; 6.
 DR SMART; SM00225; BTB; 1.
 DR SMART; SM00612; Kelch; 6.
 DR PROSITE; PS00097; BTB; 1.
 KM Cytoskeleton; Actin-binding; Kelch repeat; Repeat.
 FT DOMAIN 43 88
 FT REPEAT 212 279
 FT REPEAT 460 506
 FT REPEAT 507 553
 FT REPEAT 555 600
 FT REPEAT 601 647
 FT REPEAT 649 700
 FT REPEAT 701 747
 SEQUENCE 748 AA; 82680 MW; C11C43D8282F9F9 CRC64;

Best Local Similarity 80.0%; Pred. No. 34;
 Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;
 Qy 1 BEVVPXGMDY 10
 Db 127 BEVVP-GMDP 135
 RESULT 8
 ID PLAS_ORYZA STANDARD; PRT; 154 AA.
 AC P20423; Q9SBB8;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Plastocyanin, chloroplast precursor.
 GN PPRB.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriocaulaceae; Oryzae; Oryza.
 OC NCBI_TaxId=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=cv. Ilpoom; TISSUE=Leaf;
 RA Lee J.-S.;
 RT "Molecular cloning and characterization of plastocyanin precursor in
 rice.";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 58-154.
 RX STRAIN=cv. Japonica;
 RC MEDLINE=8936623; PubMed=2780537;
 RA Yano H., Kano M., Taguchi A., Aeo K., Nozu Y.;
 RT "The amino acid sequence of plastocyanin from rice (*Oryza sativa*,
 subspecies japonica)."
 RL Protein Seq. Data Anal. 2:385-389(1989).
 CC -1- FUNCTION: Participates in electron transfer between P700 and the
 cytochrome b6-f complex in photosystem I.
 CC -1- SUBCELLULAR LOCATION: LOOSELY BOUND TO THE INNER THYLAKOID
 MEMBRANE SURFACE IN CHLOROPLASTS.
 CC -1- SIMILARITY: Contains 1 plastocyanin-like domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF093636; AAC78108.1; -;
 DR HSSP; P00289; 2PCF.
 DR Gremlene; P20423; -;
 DR InterPro; IPR000923; BlueCu_1.
 DR InterPro; IPR001235; Copper_Blue.
 DR Pfam; PF00127; Copper-bind; 1.
 DR PRINTS; PR00156; COPPERBLU.
 DR PRODOM; PD001235; Copper_blue; 1.
 DR PROSITE; PS00196; COPPER_BLUE; 1.
 KM Chloroplast; Electron transport; Copper; Thylakoid; Membrane;
 FT TRANSIT 1 57
 FT CHAIN 58 154
 FT DOMAIN 58 154
 FT METAL 94 94
 FT METAL 139 139
 FT METAL 142 142
 FT METAL 147 147
 SEQUENCE 154 AA; 15577 MW; E45725D25B5F40D CRC64;

Query Match 63.0%; Score 34; DB 1; Length 154;
 Best Local Similarity 54.5%; Pred. No. 8.1;

Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 EEVVPKGMYS 11
 Db 100 EDVPSGVDVS 110

RESULT 9

PLAS_HORVU STANDARD; PRT; 155 AA.
 ID PLAS_HORVU
 AC P08248;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Plastocyanin, chloroplast precursor.
 GN PERE.
 OS Hordeum vulgare (Barley).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; Hordeum.
 OC NCB1_TaxID=4513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Bomi;
 RA Nielsen O.S., Gausling K.,
 RT "The precursor of barley plastocyanin: sequence of cDNA clones and
 gene expression in different tissues.",
 RL FEBS Lett. 225:159-162(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. NK 1558;
 RX MEDLINE=94039081; PubMed=8223592;
 RA Nielsen P., Gausling K.,
 RT "In vitro binding of nuclear proteins to the barley plastocyanin gene
 promoter region.",
 RL Eur. J. Biochem. 217:97-104(1993).
 CC -1- FUNCTION: Participates in electron transfer between P700 and the
 cytochrome b6-f complex in photosystem I.
 CC -1- SUBCELLULAR LOCATION: LOOSELY BOUND TO THE INNER THYLAKOID
 MEMBRANE SURFACE IN CHLOROPLASTS.
 CC -1- SIMILARITY: Contains 1 plastocyanin-like domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; Y00704; CAA68696.1; -;
 DR EMBL; Z28347; CAA82201.1; -;
 DR PIR; S38255; S38255.
 DR HSSP; P00289; 2PCF.
 DR InterPro; IPR000923; BlueCu 1.
 DR InterPro; IPR001235; Copper_blue.
 DR Pfam; PF00127; copper-bind; 1.
 DR PRINTS; PR00156; COPPERBLUB.
 DR PRODOM; PD001235; Copper_blue; 1.
 DR PROSITE; PS00196; COPPER_BLUB; 1.
 KW Chloroplast; Electron transport; Copper; Thylakoid; Membrane;
 KM TRANSIT 1
 FT CHAIN 58
 FT DOMAIN 59
 FT METAL 95
 FT METAL 140
 FT METAL 143
 FT METAL 148
 FT VARIANT 120
 SQ SEQUENCE 155 AA; 15709 MW; DAA7EABE5F6FAF91 CRC64;

Query Match 63.0%; Score 34; DB 1; Length 155;

Best Local Similarity 54.5%; Pred. No. 8.2;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 EEVVPKGMYS 11
 Db 101 EDVPSGVDVS 111

RESULT 10

PLAT_POPNI STANDARD; PRT; 168 AA.
 ID PLAT_POPNI
 AC P11970;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Plastocyanin B, chloroplast precursor.
 GN PERE.
 OS Populus nigra (lombardy poplar).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eufrosida I; Malpighiales; Salicaceae; Populus.
 OC NCB1_TaxID=3691;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Italica; TISSUE=Leaf;
 RA Reichert J., Jenzelewski V., Haehnel W.,
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 70-168.
 RC STRAIN=cv. Italica;
 RA Dimitrov M.I., Egorov C.A., Donchev A.A., Atanasov B.P.;
 RT "Complete amino acid sequence of poplar plastocyanin b.",
 RL FEBS Lett. 226:17-22(1987).
 CC -1- FUNCTION: Participates in electron transfer between P700 and the
 cytochrome b6-f complex in photosystem I.
 CC -1- SUBCELLULAR LOCATION: LOOSELY BOUND TO THE INNER THYLAKOID
 MEMBRANE SURFACE IN CHLOROPLASTS.
 CC -1- SIMILARITY: THERE ARE 12 DIFFERENCES BETWEEN THE SEQUENCE OF
 POPULAR PLASTOCYANINS A AND B.
 CC -1- SIMILARITY: Contains 1 plastocyanin-like domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; Z50186; CAA90565.1; -;
 DR PIR; S00210; S00210.
 DR PIR; S58208; S58208.
 DR HSSP; P00299; 1PLC.
 DR InterPro; IPR000923; BlueCu 1.
 DR InterPro; IPR001235; Copper_blue.
 DR Pfam; PF00127; copper-bind; 1.
 DR PRINTS; PR00156; COPPERBLUB.
 DR PRODOM; PD001235; Copper_blue; 1.
 DR PROSITE; PS00196; COPPER_BLUB; 1.
 KW Chloroplast; Electron transport; Copper; Thylakoid; Membrane;
 KM TRANSIT 69
 FT CHAIN 70
 FT DOMAIN 70
 FT METAL 106
 FT METAL 153
 FT METAL 156
 FT METAL 161
 SQ SEQUENCE 168 AA; 16981 MW; F20DA6EA2038AEEA CRC64;

Query Match 63.0%; Score 34; DB 1; Length 168;
 Best Local Similarity 54.5%; Pred. No. 8.9;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

QY      1 EEVVPXGMDYS 11
DB      112 EDVAVPGCVDS 122

RESULT 11
SYFB BORBU      STANDARD;      PRT;      566 AA.
ID P94283;
AC 30-MAY-2000 (rel. 39, Created)
DT 30-MAY-2000 (rel. 39, Last sequence update)
DE 28-FEB-2003 (rel. 41, Last annotation update)
DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--
DE tRNA ligase beta chain) (PHERS).
OS PHER OR BB0514;
OC Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxId=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RA Barbour A.G., Hinnebusch J.;
RT "Phenylalanyl-tRNA synthetase genes (alpha and beta subunits) and
RT thioredoxin reductase gene of Borrelia burgdorferi.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Frazer C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty J., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA Van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Uterback T., Matthey L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi.";
RL Nature 390:580-586(1997).
CC -1- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
CC dihydrophosphate + L-phenylalanyl-tRNA(Phe).
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE PHENYLALANYL-TRNA SYNTHETASE BETA
CC CHAIN FAMILY. SUBFAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U82978; AAB41019.1; -
DR EMBL; AE001153; AAC66870.1; -
DR PIR; A70164; A70164.
DR TIGR; BB0514; -
DR HANAP; MF_00284; -; 1.
DR InterPro; IPR005147; B5.
DR InterPro; IPR004531; Pher_arch.
DR Pfam; PF03484; B5; 1.
DR TIGRFAMs; TIGR00471; phet_arch; 1.
KW Aminoacyl-tRNA synthetase, Protein biosynthesis, ligase, ATP-binding;
KW Complete proteome.
SQ SEQUENCE 566 AA; 65173 MW; 9D48CB5D6D3B74B CRC64;
Query Match 63.0%; Score 34; DB 1; Length 566;
Best Local Similarity 85.7%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      4 VEPXGMDY 10
DB      169 VEPXGMDY 175

RESULT 12
RBL4 ARATH
ID RBL4 ARATH      STANDARD;      PRT;      908 AA.
AC Q9F0K8;
DT 15-SEP-2003 (rel. 42, Created)
DT 15-SEP-2003 (rel. 42, Last sequence update)
DE 15-SEP-2003 (rel. 42, Last annotation update)
DE Probable disease resistance RPP8-like protein 4.
GN RPP8L4 OR AT5G48620 OR K1SN18.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Burkholderia; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxId=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=99087489; PubMed=9872454;
RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
RT sequence features of the regions of 1,013,767 bp covered by sixteen
RT physically assigned P1 and TAC clones.";
RL DNA Res. 5:297-308(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA Hayashizaki Y., Shinozaki K.;
RT "Arabidopsis thaliana full-length cDNA.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Potential disease resistance protein.
CC -1- DOMAIN: The LRR repeats probably act as specificity determinant of
CC pathogen recognition (By similarity).
CC -1- SIMILARITY: Belongs to the disease resistance NB-LRR family.
CC RPP8/LRR subfamily.
CC -1- SIMILARITY: Contains 3 leucine-rich (LRR) repeats.
CC -1- SIMILARITY: Contains 1 NB-ARC domain.
CC -1- DATABASE: NAME=NIB-LRRs;
CC NOTE=Functional and comparative genomics of disease resistance gene
CC homologs;
CC WWW="http://niblrr.ucdavis.edu".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB015468; BAB10635.1; -
DR EMBL; AK117163; BAC11841.1; -
DR InterPro; IPR000767; Disease_resist.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR002182; NB-ARC.
DR Pfam; PF00560; LRR; 2.
DR Pfam; PF00931; NB-ARC; 1.
DR PRINTS; PR00364; DISBASERIST.
KW Plant defense, ATP-binding, Repeat, Leucine-rich repeat.
KW DOMAIN 10 45
KW DOMAIN 146 459
KW REPEAT 575 599
KW REPEAT 575 599
KW REPEAT 600 623
KW REPEAT 842 867
KW NP_BIND 192 199
KW ATP (POTENTIAL).

```

SQ	SEQUENCE	908 AA;	104448 MW;	3111991817239693 CRC64;
	Query Match		63.0%;	Score 34; DB 1; Length 908;
	Best Local Similarity		60.0%;	Pred. No. 53;
	Matches	6; Conservative	2; Mismatches	2; Indels 0; Gaps
Oy		1 EEWVPGMDY 10		
		::::		
Db		883 EKLVPGEEDY 892		

RESULT 13			
ID	RPP8_ARATH	STANDARD;	PRT; 908 AA.
AC	Q8W4J9; Q8GMG5; Q9M5A1; Q9ZSY3; Q9ZSY4;		
DT	15-SEP-2003 (Rel. 42, Created)		
DT	15-SEP-2003 (Rel. 42, Last sequence update)		
DE	15-SEP-2003 (Rel. 42, Last annotation update)		
DR	Disease resistance protein RPP8 (Resistance to Peronospora parasitica protein 8).		
GN	RPP8 OR HRT OR ATSG43470 OR MRP20.19.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
CC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eucotids II; Brassicales; Brassicaceae; Arabidopsids.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, MUTANTS RPP8-1, RPP8-2 AND		
RP	RPP8-3, AND VARIANTS.		
RC	STRAIN=cv. Columbia, and cv. Landsberg erecta;		
RX	MEDLINE=9030193; PubMed=9811794;		
RA	McDowell J.M., Dhandaaydam M., Long T.A., Aarts M.G.M., Goff S.,		
RA	Holtub E.B., Dangl J.L.;		
RT	"Intragenetic recombination and diversifying selection contribute to the		
RT	evolution of downy mildew resistance at the RPP8 locus of		
RT	Arabidopsis."		
RL	Plant Cell 10:1861-1874(1998).		
RN	[2]		
RP	SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS.		
RP	STRAIN=cv. Di-17;		
RX	MEDLINE=20271766; PubMed=10810142;		
RA	Cosley M.B., Pathirana S., Wu H.-J., Kachroo P., Kleisig D.F.;		
RT	"Members of the Arabidopsis HRT/RPP8 family of resistance genes confer		
RT	resistance to both viral and oomycete pathogens."		
RL	Plant Cell 12:663-676(2000).		
RN	[3]		
RP	SEQUENCE FROM N.A. (ISOFORM 1).		
RC	STRAIN=cv. Columbia;		
RX	MEDLINE=20181125; PubMed=10718197;		
RA	Sato S., Nakamura Y., Kaneko T., Katoh T., Asanizu E., Kotani H.,		
RA	Tabata S.;		
RT	"Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence		
RT	features of the regions of 3,076,755 bp covered by sixty P1 and TAC		
RT	clones.";		
RL	DNA Res. 7:31-63(2000).		
RN	[4]		
RP	SEQUENCE FROM N.A. (ISOFORM 1).		
RC	STRAIN=cv. Columbia;		
RA	Shinozaki K., Davis R.W., Eckert J.R., Theologis A.;		
RT	"RIKEN Arabidopsis full length cDNA clones (RAFTS) sequenced by the		
RT	SSP consortium (Saik/Stamford/PESC)."		
RL	Submitted (NOV-2001) to the EMBL/GenBank/DDB databases.		
RN	[5]		
RP	SEQUENCE FROM N.A. (ISOFORM 2).		
RC	STRAIN=cv. Columbia;		
RA	Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,		
RA	Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,		
RA	Hayashizaki Y., Shinozaki K.;		
RT	"Arabidopsis thaliana full-length cDNA.";		
RL	Submitted (NOV-2002) to the EMBL/GenBank/DDB databases.		
RN	[6]		
RP	INTERACTION WITH TIP.		
RX	MEDLINE=20496823; PubMed=11041886;		

RA	Ren T., Yu F., Morris T.J.:
RT	"HRT gene function requires interaction between a NAC protein and
RL	viral capsid protein to confer resistance to turnip crinkle virus.";
CC	Plant Cell 12:1917-1926(2000).
CC	-1- FUNCTION: Disease resistance protein. Resistance proteins guard
CC	the plant against pathogens that contain an appropriate avirulence
CC	protein via an indirect interaction with this avirulence protein.
CC	That triggers a defense system including the hypersensitive
CC	response, which restricts the pathogen growth. The interaction
CC	with TIR (TIRV-interacting protein) may be essential for the
CC	recognition of the avirulence proteins, and the triggering of the
CC	defense response.
CC	-1- SUBUNIT: Interacts with the NAC protein TIRP.
CC	-1- ALTERNATIVE PRODUCTS:
CC	Event=Alternative splicing; Named isoforms=2;
CC	Name=1;
CC	IsoId=Q8W4U9-1; Sequence=Displayed;
CC	Name=2;
CC	IsoId=Q8W4U9-2; Sequence=VSP_007171, VSP_007172;
CC	Note=Has been shown to exist "only" in cv. Columbia so far;
CC	-1- DOMAIN: The LRR repeats probably act as specificity determinant of
CC	pathogen recognition.
CC	-1- POLYMORPHISM: The strong polymorphisms present in cv. Di-17 and
CC	cv. Columbia are probably due to an unequal crossing-over between
CC	the highly related RPP8 and RPPHA genes present in cv. Landsberg
CC	erecta. Such variations probably modify the specificity of
CC	pathogen recognition.
CC	-1- MISCELLANEOUS: In cv. Columbia and cv. Landsberg erecta, RPP8
CC	specifically recognizes the Emco5 avirulence protein from
CC	Pemonespora parasitica, while it is not the case in cv. Di-17,
CC	where it confers resistance to Turnip Crinkle Virus upon
CC	recognition of the viral capsid protein.
CC	-1- SIMILARITY: Belongs to the disease resistance NB-LRR family.
CC	RPP8/HRT subfamily.
CC	-1- SIMILARITY: Contains 2 leucine-rich (LRR) repeats.
CC	-1- SIMILARITY: Contains 1 NB-ARC domain.
CC	-1- DATABASE: NAME=NIB-LRRS;
CC	NOTE=Functional and comparative genomics of disease resistance gene
CC	homologs;
CC	WWW="http://nblrrs.ucdavis.edu".
CC	-----
CC	THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb.ch/announce/
CC	or send an email to license@isb.ch).
CC	-----
DR	EMBL; AF089710; AAC83165.1; -
DR	EMBL; AF089711; AAC78631.1; -
DR	EMBL; AF234174; AAF36987.1; -
DR	EMBL; AB025638; BAA97426.1; -
DR	EMBL; AY062514; AAL32592.1; -
DR	EMBL; AK118862; BAC43449.1; -
DR	InterPro; IPR00767; Disease_resist.
DR	InterPro; IPR001611; LRR.
DR	InterPro; IPR002182; NB-ARC.
DR	Pfam; PF00560; LRR. 2.
DR	Pfam; PF00931; NB-ARC. 1.
DR	PRINTS; PR00364; DISEASERST
KM	Plant defense; ATP-binding; Repeat; Leucine-rich repeat;
KW	Alternative splicing; Polymorphism.
FT	DOMAIN 10 45
FT	NB-ARC. LEUCINE-ZIPPER.
FT	REPEAT 146 459
FT	LRR 1. LRR 1.
FT	REPAT 842 867
FT	LRR 2. LRR 2.
FT	NP_BIND 192 199
FT	ATP (POTENTIAL).
FT	VARSPLIC 294 308
FT	WKMLLTRNSGVGTH -> ELLMYIHEALFLNLS (in
FT	isoform 2).
FT	/FtId=VSP_007171.
FT	Mising (in isoform 2).
FT	/FtId=VSP_007172.
FT	VARSPLIC 309 908

```

FT  VARIANT 27 31 IDGOL -> VDEOI (in cv. Landsberg erecta).
FT  VARIANT 29 29 G -> B (in cv. Di-17).
FT  VARIANT 85 87 SGK -> RGE (in cv. Di-17 and cv.
FT  VARIANT 88 91 GKNV -> EKG (in cv. Landsberg erecta).
FT  VARIANT 100 100 C -> R (in cv. Landsberg erecta).
FT  VARIANT 121 121 E -> D (in cv. Landsberg erecta).
FT  VARIANT 129 129 F -> L (in cv. Landsberg erecta).
FT  VARIANT 133 133 Q -> OQ (in cv. Di-17).
FT  VARIANT 138 138 G -> V (in cv. Landsberg erecta).
FT  VARIANT 173 173 K -> T (in cv. Di-17).
FT  VARIANT 177 177 G -> C (in cv. Di-17).
FT  VARIANT 253 253 Y -> S (in cv. Di-17).
FT  VARIANT 257 257 R -> P (in cv. Di-17).
FT  VARIANT 265 265 A -> T (in cv. Di-17).
FT  VARIANT 267 267 R -> K (in cv. Landsberg erecta).
FT  VARIANT 270 270 V -> L (in cv. Di-17).
FT  VARIANT 283 283 A -> R (in cv. Di-17).
FT  VARIANT 369 369 V -> V (in cv. Di-17 and cv. Landsberg
FT  VARIANT 387 387 erecta).
FT  VARIANT 399 399 F -> S (in cv. Di-17 and cv. Landsberg
FT  VARIANT 426 426 W -> C (in cv. Di-17 and cv. Landsberg
FT  VARIANT 429 429 erecta).
FT  VARIANT 429 429 C -> R (in cv. Di-17 and cv. Landsberg
FT  VARIANT 429 429 erecta).
FT  VARIANT 436 446 N -> F (in cv. Di-17 and cv. Landsberg
FT  VARIANT 438 443 DSEISTSLFY -> YSKISAYDLFN (in cv.
FT  VARIANT 450 443 Landsberg erecta).
FT  VARIANT 460 450 EISTYS -> KITTOE (in cv. Di-17).
FT  VARIANT 480 460 A -> Q (in cv. Landsberg erecta).
FT  VARIANT 481 486 B -> V (in cv. Landsberg erecta).
FT  VARIANT 485 481 DNLSMQ -> NKYLKRVH (in cv. Di-17).
FT  VARIANT 489 486 DN -> NR (in cv. Landsberg erecta).
FT  VARIANT 514 489 MQ -> SH (in cv. Landsberg erecta).
FT  VARIANT 518 514 Y -> N (in cv. Landsberg erecta).
FT  VARIANT 518 518 I -> K (in cv. Di-17 and cv. Landsberg
FT  VARIANT 519 518 erecta).
FT  VARIANT 519 518 C -> S (in cv. Di-17 and cv. Landsberg
FT  VARIANT 528 528 erecta).
FT  VARIANT 537 537 T -> I (in cv. Di-17).
FT  VARIANT 542 537 S -> R (in cv. Di-17).
FT  VARIANT 543 542 H -> Q (in cv. Di-17).
FT  VARIANT 549 543 I -> L (in cv. Landsberg erecta).
FT  VARIANT 549 550 KNKT -> RNNA (in cv. Di-17).
FT  VARIANT 558 549 K -> N (in cv. Landsberg erecta).
FT  VARIANT 558 565 PRFEDYW -> WDEDFG (in cv. Landsberg
FT  VARIANT 558 560 erecta).
FT  VARIANT 564 564 PRF -> SRFK (in cv. Di-17).
FT  VARIANT 584 564 Y -> F (in cv. Di-17).
FT  VARIANT 595 595 S -> Y (in cv. Landsberg erecta).
FT  VARIANT 595 595 C -> S (in cv. Di-17 and cv. Landsberg
FT  VARIANT 609 612 erecta).
FT  VARIANT 630 612 YEAK -> FLAG (in cv. Landsberg erecta).
FT  VARIANT 632 635 NLRVDT -> DLSTVH (in cv. Di-17).
FT  VARIANT 650 636 RVDTR -> SVNNK (in cv. Landsberg erecta).
FT  VARIANT 653 650 Q -> E (in cv. Di-17).
FT  VARIANT 659 653 YL -> HI (in cv. Di-17).
FT  VARIANT 660 660 M -> K (in cv. Di-17).
FT  VARIANT 676 660 Missing (in cv. Di-17).
FT  VARIANT 678 678 M -> K (in cv. Landsberg erecta).
FT  VARIANT 692 678 YLY -> FLV (in cv. Landsberg erecta).
FT  VARIANT 697 692 YG -> FR (in cv. Di-17).
FT  VARIANT 700 697 R -> H (in cv. Landsberg erecta).
FT  VARIANT 700 700 R -> Q (in cv. Di-17).
FT  VARIANT 700 700 A -> G (in cv. Di-17).

```

Query Match 63.0%; Score 34; DB 1; Length 908;
 Best Local Similarity 60.0%; Pred. No. 53;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMDY 10

```

DB 883 EKLVPGEEDY 892
RESULT 14
RPH8 ARATH STANDARD; PRT; 910 AA.
AC P59584;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Disease resistance protein RPH8A (RPH8 homolog A).
GN RPH8A.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SOURCE FROM N.A., AND FUNCTION.
RC STRAIN=cv. Landsberg erecta;
RX MEDLINE=99030193; PubMed=9811794;
RA McDowell J.M., Dhandaydham M., Long T.A., Aarts M.G.N., Goff S.,
RA Holub E.B., Dangl J.L.;
RT "Intergenic recombination and diversifying selection contribute to the
RT evolution of downy mildew resistance at the RPP8 locus of
RT Arabidopsis."
RL Plant Cell 10:1861-1874 (1998).
CC -1- FUNCTION: Disease resistance protein. Resistance proteins guard
CC the plant against pathogens that contain an appropriate avirulence
CC protein via an indirect interaction with this avirulence protein.
CC That triggers a defense system including the hypersensitive
CC response, which restricts the pathogen growth. In contrast to
CC RPP8, it does not specifically recognize the Emco5 avirulence
CC protein from Peronospora parasitica.
CC -1- MISCELLANEOUS: In cv. Columbia and cv. Di-17, this protein is not
CC present due to an unequal crossing over between the RPP8 and RPH8A
CC genes that creates a unique RPP8 gene.
CC -1- SIMILARITY: Belongs to the disease resistance NB-LRR family.
CC RPP8/HRT subfamily.
CC -1- SIMILARITY: Contains 2 leucine-rich (LRR) repeats.
CC -1- SIMILARITY: Contains 1 NB-ARC domain.
CC -1- DATABASE: NAME=NIB-LRRS;
CC NOTE=Functional and comparative genomics of disease resistance gene
CC homologs;
CC WWW="http://niblr.ucdavis.edu".
KW Plant defense; ATP-binding; Repeat; leucine-rich repeat.
FT DOMAIN 10 45 LEUCINE-ZIPPER.
FT DOMAIN 146 459 NB-ARC.
FT REPEAT 602 625 LRR 1.
FT REPEAT 844 869 LRR 2.
FT NP BIND 192 199 ATP (POTENTIAL).
SQ SEQUENCE 910 AA; 105263 MW; 5B1B9F5A19A12EB CRC64;

```

Query Match 63.0%; Score 34; DB 1; Length 910;
 Best Local Similarity 60.0%; Pred. No. 53;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMDY 10

Db 885 EKLVPGEEDY 894

RESULT 15
 Y939_METUA STANDARD; PRT; 276 AA.
 ID Y939_METUA
 AC Q58349;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein M0939.
 GN M0939.
 OS Mechanococcus jannaschii.

```

OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=9637999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weldman J.F., Fuhrmann J.L., Nguyen D.,
RA Utecherback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67537; AAB98946.1; -.
DR PIR; C64417; C64417.
DR TIGR; MJ0939; -.
DR Pfam; PF03692; UPP0153; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 276 AA; 33454 MW; 97BD69D392BC8FDF CRC64;
OY 1 EEVVPXGMDYS 11
Db 141 EELIENGMEHS 151

```

```

Query Match 61.1%; Score 33; DB 1; Length 276;
Best Local Similarity 45.5%; Pred. No. 24;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```

Search completed: December 22, 2003, 17:42:26
 Job time : 6.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:27:26 ; Search time 25.2 Seconds

(without alignments)
112.642 Million cell updates/sec

Title: US-09-909-164-8

Sequence: 1 EEVVPXGMDYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_ROOT:*
12: SP_VIRUS:*
13: SP_VIRTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	40	74.1	156 3	Q12479
2	38	70.4	319 16	Q8ESV7
3	38	70.4	363 17	Q30260
4	38	70.4	1044 16	Q8DIH0
5	36	66.7	341 10	Q22081
6	36	66.7	348 10	Q22096
7	36	66.7	452 10	Q8W568
8	36	66.7	460 10	Q8C9T7
9	36	66.7	1047 10	P93782
10	36	66.7	1083 10	Q9SN30
11	36	66.7	1084 10	Q43010
12	36	66.7	1100 10	Q8S064
13	35	64.8	219 5	Q9GQ04
14	35	64.8	253 16	Q8XP48
15	35	64.8	298 2	O52367
16	35	64.8	368 2	Q8KL43

17	35	64.8	425 5	Q9XVK4	Q9XVK4 caenorhabdi
18	35	64.8	433 16	Q9A382	Q9A382 caulobacter
19	35	64.8	440 17	Q9YR13	Q9YR13 aeropyrum p
20	35	64.8	511 2	O52680	O52680 escherichia
21	35	64.8	517 16	Q8X2L5	Q8X2L5 ralslonia s
22	35	64.8	595 10	Q946J7	Q946J7 androgaphi
23	35	64.8	745 5	Q9SP46	Q9SP46 carcinos ma
24	35	64.8	1031 5	Q9U6A3	Q9U6A3 callinectes
25	35	64.8	1150 5	O17704	O17704 caenorhabdi
26	35	64.8	1410 2	O52673	O52673 escherichia
27	35	64.8	1420 2	O52666	O52666 escherichia
28	35	64.8	1474 17	O27146	O27146 methanobact
29	35	64.8	1828 16	Q98KZ9	Q98KZ9 rhizobium l
30	34.5	63.9	748 4	Q8TBJ7	Q8TBJ7 homo sapien
31	34	63.0	143 5	Q9VSY8	Q9VSY8 drosophi
32	34	63.0	215 16	Q8R3L5	Q8R3L5 thermoaer
33	34	63.0	222 16	Q8E5L9	Q8E5L9 streptococ
34	34	63.0	222 16	Q8DZW9	Q8DZW9 streptococ
35	34	63.0	284 2	O8GP33	O8GP33 lactobacill
36	34	63.0	290 16	Q8U7J0	Q8U7J0 agrobacteri
37	34	63.0	296 17	Q9YET8	Q9YET8 aeropyrum p
38	34	63.0	357 17	O29920	O29920 archaeoglob
39	34	63.0	366 17	O29451	O29451 archaeoglob
40	34	63.0	387 16	Q98FX1	Q98FX1 rhizobium l
41	34	63.0	543 3	O8TFP4	O8TFP4 trichoderma
42	34	63.0	558 16	Q8R8Z2	Q8R8Z2 thermoaer
43	34	63.0	565 16	Q9CINI	Q9CINI lactococcus
44	34	63.0	587 16	Q9JZP8	Q9JZP8 neisseria m
45	34	63.0	671 16	Q8G3J2	Q8G3J2 bifidobacte

ALIGNMENTS

RESULT 1
ID Q12479 PRELIMINARY, PRT, 156 AA.
AC Q12479;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ORF YOR013W.
GN YOR013W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP De haan M., Grivell L.A., Marsee A.C.;
RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP MIPs;
RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP STRAIN=FY1679;
RA De haan M., Marsee A.C., Grivell L.A.;
RA Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP STRAIN=FROM N.A.
RC STRAIN=FY1679;
RX MEDLINE=94019318; PubMed=8413243;
RA Dumont M.E., Schlichter J.B., Cardillo T.S., Hayes M.K., Bethlenny G.,
RA Sherman P.;
RT "CYC2 encodes a factor involved in mitochondrial import of yeast
RT cytochrome c.";
RL Mol. Cell. Biol. 13:6442-6451(1993).
RN [5]
RP STRAIN=FY1679;
RC STRAIN=FROM N.A.
RX MEDLINE=94169519; PubMed=7764548;

RA Lee Y.S., Shimizu J., Yoda K., Yamasaki M.,
 RT "Molecular cloning of a gene, DHS1, which complements a drug-
 RT hypersensitive mutation of the yeast *Saccharomyces cerevisiae*,"
 RL BioSci. Biotechnol. Biochem. 58:391-395(1994).
 DR EMBL; 274920; CAA99201.1; -.
 DR EMBL; X87331; CAA60762.1; -.
 DR SCD; S0005539; YOR013W.
 SQ SEQUENCE 156 AA; 17881 MW; 380442B74C272B41 CRC64;
 Query Match 74.1%; Score 40; DB 3; Length 156;
 Best Local Similarity 77.8%; Pred. No. 2.3;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 EVVXPXGMDY 10
 DB 50 EVMPLGMDY 58
 RESULT 2
 Q8ESV7 PRELIMINARY; PRT; 319 AA.
 ID Q8ESV7;
 AC Q8ESV7;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Hypothetical conserved protein.
 GN O80509.
 OS Oceanobacillus thelyensis.
 OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.
 NCBI_TaxId=162710;
 RX NCBI_TaxId=162710;
 RP SEQUENCE FROM N.A.
 RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
 RX MEDLINE=22220767; PubMed=12235376;
 RA Takami H., Takaki Y., Uchiyama I.;
 RT "Genome sequence of Oceanobacillus thelyensis isolated from the Iheya
 RT Ridge and its unexpected adaptive capabilities to extreme
 RT environments";
 RL Nucleic Acids Res. 30:3927-3935(2002).
 DR EMBL; AP004594; BAC12465.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 319 AA; 35617 MW; 3BDABAF13E79E37 CRC64;
 Query Match 70.4%; Score 38; DB 16; Length 319;
 Best Local Similarity 60.0%; Pred. No. 14;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EEVVPXGMDY 10
 DB 189 EQLVPHGIDY 198
 RESULT 3
 O30260 PRELIMINARY; PRT; 363 AA.
 ID O30260;
 AC O30260;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Hypothetical protein AF2411.
 GN AF2411.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
 OC Archaeoglobaceae; Archaeoglobus.
 NCBI_TaxId=2234;
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.J., Gwin M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrleides N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,

RA Kirtrees E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodok A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.P., McDonald L., Utterback T.,
 RA Cotton M.D., Spitzig T., Artlich P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Maon T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon *Archaeoglobus fulgidus*,"
 RL Nature 390:364-370(1997).
 DR EMBL; AE001109; AAB91255.1; -.
 DR TIGR; AF2411; -.
 DR InterPro; IPR002103; Bac_luciferase.
 DR Pfam; PF00296; bac_luciferase; 1..
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 363 AA; 41736 MW; 0E976EAE788F4803 CRC64;
 Query Match 70.4%; Score 38; DB 17; Length 363;
 Best Local Similarity 54.5%; Pred. No. 16;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 EEVVPXGMDYS 11
 DB 120 ENIVPYGIDFS 130
 RESULT 4
 Q8DIH0 PRELIMINARY; PRT; 1044 AA.
 ID Q8DIH0
 AC Q8DIH0;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Multidrug efflux transporter.
 GN TL1618.
 OS *Synechococcus elongatus* (Thermosynechococcus elongatus).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
 NCBI_TaxId=32046;
 RX NCBI_TaxId=32046;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP-1;
 RX MEDLINE=22225144; PubMed=12240834;
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 RA Matanabe A., Iriyuchi M., Kawashima K., Kimura T., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
 RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the thermophilic cyanobacterium
 RT *Thermosynechococcus elongatus* BP-1,"
 RL DNA Res. 9:123-130(2002).
 DR EMBL; AP005374; BAC09170.1; -.
 KW Complete proteome.
 SQ SEQUENCE 1044 AA; 113205 MW; 00E9C13F0F636D2F CRC64;
 Query Match 70.4%; Score 38; DB 16; Length 1044;
 Best Local Similarity 63.6%; Pred. No. 53;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 EEVVPXGMDYS 11
 DB 843 EEVLPNGICYS 853
 RESULT 5
 O22081 PRELIMINARY; PRT; 341 AA.
 ID O22081
 AC O22081;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Sucrose-phosphate synthase (Fragment).
 GN CITSP52.
 OS Citrus unshiu (Satsuma orange).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Sapindales; Rutaceae; Citrus.
OX NCBI_TaxID=55188;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Miyagawa-Wase; TISSUE=juice sacs and segment epidermis;
RX MEDLINE=96439842; PubMed=8842155;
RA Komatsu A., Takanoakura Y., Omura M., Akihama T.;
RT "Cloning and molecular analysis of cDNA encoding three sucrose
RT phosphate synthase isoforms from a citrus fruit (Citrus unshiu
RT Marc.).";
RL Mol. Genet. 252:346-351 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Miyagawa-Wase; TISSUE=juice sacs and segment epidermis;
RA Komatsu A., Takanoakura Y., Moriguchi T., Omura M., Akihama T.;
RT "Differential expression of three sucrose-phosphate synthase isoforms
RT during sucrose accumulation in citrus fruits (Citrus unshiu Marc.).";
RL Plant Sci. 140:169-178 (1999).
DR EMBL; AB006319; BAA23215.1; -.
FT NON_TER 1 341
FT SEQUENCE 341 AA; 38136 MW; 61417A69C4560777 CRC64;
SQ
Query Match 66.7%; Score 36; DB 10; Length 341;
Best Local Similarity 66.7%; Pred. No. 40;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
OY 3 VVPXGMDYS 11
DB 228 VPPGMDPS 236
RESULT 6
ID O22096 PRELIMINARY; PRT; 348 AA.
AC O22096;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Sucrose-phosphate synthase (Fragment).
GN CITSP83.
OS Citrus unshiu (Satsuma orange).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Sapindales; Rutaceae; Citrus.
OX NCBI_TaxID=55188;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=juice sacs and segment epidermis;
RX MEDLINE=96439842; PubMed=8842155;
RA Komatsu A., Takanoakura Y., Omura M., Akihama T.;
RT "Cloning and molecular analysis of cDNA encoding three sucrose
RT phosphate synthase isoforms from a citrus fruit (Citrus unshiu
RT Marc.).";
RL Mol. Genet. 252:346-351 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=juice sacs and segment epidermis;
RA Komatsu A., Takanoakura Y., Moriguchi T., Omura M., Akihama T.;
RT "Differential expression of three sucrose-phosphate synthase isoforms
RT during sucrose accumulation in citrus fruits (Citrus unshiu Marc.).";
RL Plant Sci. 140:169-178 (1999).
DR EMBL; AB006660; BAA22071.1; -.
FT NON_TER 1 348
FT SEQUENCE 348 AA; 38556 MW; EB1C21EBA6F5C5E CRC64;
SQ
Query Match 66.7%; Score 36; DB 10; Length 348;
Best Local Similarity 66.7%; Pred. No. 41;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
OY 3 VVPXGMDYS 11

```

```

DB 234 VPPGMDPS 242
RESULT 7
ID Q8W568 PRELIMINARY; PRT; 452 AA.
AC Q8W568;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE At5g3750/F25P22.17.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bower L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Kim C.J., Chen H., Cheuk R., Shin P., Banh J., Bower L.,
RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF419606; AAL1938.1; -.
DR EMBL; AY113044; AAM47352.1; -.
DR InterPro; IPR002471; Prol endopep ser.
DR InterPro; IPR000379; Ser estere site.
DR PROSITE; PS00709; PRO_ENDOPEP_SER; 1.
SQ SEQUENCE 452 AA; 49682 MW; A15995B21742C4A CRC64;
Query Match 66.7%; Score 36; DB 10; Length 452;
Best Local Similarity 70.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 EEVVPXGMDY 10
DB 210 EEDVPSAMDY 219
RESULT 8
ID Q9C9T7 PRELIMINARY; PRT; 460 AA.
AC Q9C9T7;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical 50.6 KDa protein.
GN F25P22.17.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;

```

RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federpiel N.A., Kaul S.,
 White O., Alonso J., Altali H., Araujo R., Bowman C.L., Brooks S.Y.,
 Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 Dunn P., Etlug P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 Gull J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 Kim C.J., Koo H.U., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marzilli A.,
 Miltescher J., Miranda M., Niernan M.C., Osborne B.I.,
 Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.W.,
 Sun H., Tallon L.J., Tambung G., Toriumi M.J., Town C.D.,
 Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RA "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana";
 RT
 RL Nature 408:816-820(2000).
 DR EMBL; AC012679; AAG52073.1; -;
 DR InterPro; IPR002471; Prol_endopep_ser.
 DR InterPro; IPR000379; Ser_estre site.
 DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 460 AA; 50564 MW; E94B27B5C4B249EC CRC64;

Query Match 66.7%; Score 36; DB 10; Length 460;
 Best Local Similarity 70.0%; Pred. No. 56;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EVVFXGMDY 10
 Db 218 EEDVPSAMDY 227

RESULT 9
 P93782 PRELIMINARY; PRT; 1047 AA.
 AC P93782;
 DT 01-MAY-1997 (TRENBLREL. 03, Created)
 DT 01-MAY-1997 (TRENBLREL. 03, Last sequence update)
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE Sucrose-phosphate synthase (EC 2.4.1.14) (Fragment).
 GN SOSPS1.
 OS Saccharum officinarum (Sugarcane).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC PACCAD clade; Panicoideae; Andropogoneae; Saccharum.
 OK NCBI_TaxID=4547;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=leaf;
 RA Sugiharto B., Sakakibara H., Sugiyama T.;
 RT "Differential Expression of Two Genes for Sucrose-Phosphate Synthase
 in Sugarcane: Molecular Cloning of the cDNAs and Comparative Analysis
 of Gene Expression";
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB001337; BAA19241.1; -;
 DR InterPro; IPR001296; Glyco_transf_1.
 DR Pfam; PF00534; Glycosyltransferase; Transf_1.
 KW Glycosyltransferase; Transferase.
 FT NON_TER
 SQ SEQUENCE 1047 AA; 116379 MW; D0EDB34961E1D83D CRC64;

Query Match 66.7%; Score 36; DB 10; Length 1047;
 Best Local Similarity 66.7%; Pred. No. 14e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDY 11
 Db 414 VIPPGMDFS 422

RESULT 10
 Q9SN30 PRELIMINARY; PRT; 1083 AA.
 AC Q9SN30;
 DT 01-MAY-2000 (TRENBLREL. 13, Created)
 DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE Sucrose-phosphate synthase-like protein (EC 2.4.1.14).
 GN F2EM11.40 OR AtG10120.
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC Eucoside II; Brassicales; Brassicaceae; Arabidopsis.
 OK NCBI_TaxID=3702;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Beyer M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
 May K.F.X., Lemcke K., Scheller C.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL049487; CAB39764.1; -;
 DR EMBL; AL161516; CAB78135.1; -;
 DR InterPro; IPR001296; Glyco_transf_1.
 DR Pfam; PF00534; Glycosyltransferase; Transf_1.
 KW Glycosyltransferase; Transferase.
 SQ SEQUENCE 1083 AA; 122688 MW; EFC126FCA2137B1 CRC64;

Query Match 66.7%; Score 36; DB 10; Length 1083;
 Best Local Similarity 66.7%; Pred. No. 1.5e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDY 11
 Db 483 VIPPGMDFS 491

RESULT 11
 Q43010 PRELIMINARY; PRT; 1084 AA.
 AC Q43010;
 DT 01-NOV-1996 (TRENBLREL. 01, Created)
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE Sucrose phosphate synthase.
 GN SPS1.
 OS Oryza sativa (Rice).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC Euphorbiaceae; Oryzaeae; Oryza.
 OK NCBI_TaxID=4530;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. japonica;
 RA Sakamoto M., Satozawa T., Kishimoto N., Higo K., Shinada H.,
 Fujimura T.;
 RT "Structure and RFLP mapping of a rice sucrose phosphate synthase (SPS)
 gene that is specifically expressed in the source organ";
 RL Plant Sci. 112:207-217(1995).
 DR EMBL; D45890; BAA08304.1; -;
 DR Gramene; Q43010; -;
 DR InterPro; IPR001296; Glyco_transf_1.
 DR InterPro; IPR001557; L_LDH.
 DR Pfam; PF00534; Glycosyltransferase; Transf_1; 1.
 DR PROSITE; PS00064; L_LDH; 1.
 SQ SEQUENCE 1084 AA; 119417 MW; 4D0D79AFEDD2F92B CRC64;

Query Match 66.7%; Score 36; DB 10; Length 1084;
 Best Local Similarity 66.7%; Pred. No. 1.5e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 VVPGXGMDYS 11
 |||
 Db 453 VVPGXGMDYS 461

RESULT 12

O8S064 PRELIMINARY; PRT; 1100 AA.

AC O8S064; ID O8S064; PRELIMINARY; PRT; 1100 AA.
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Putative sucrose-phosphate synthase.
 GN P0678F11.14.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriocaridaceae; Oryzae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nippondare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nippondare (Ga3) genomic DNA, chromosome 1, PAC
 clone:p0678F11."
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP003437; BAB6107.1; -
 DR O8S064; -
 DR Gremene; O8S064; -
 DR InterPro; IPR001296; Glyco_trans_1.
 DR InterPro; IPR001557; L_LDH.
 DR Pfam; PF00534; Glycos_transf_1; 1.
 DR PROSITE; PS00064; L_LDH; 1.
 SQ SEQUENCE 1100 AA; 121170 MW; 084F4604BA389CAD CRC64;

Query Match 66.7%; Score 36; DB 10; Length 1100;
 Best Local Similarity 66.7%; Pred. No. 1.5e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 VVPGXGMDYS 11
 |||
 Db 469 VVPGXGMDYS 477

RESULT 13

O9G004 PRELIMINARY; PRT; 219 AA.

AC O9G004; ID O9G004; PRELIMINARY; PRT; 219 AA.
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE Na+/K+/2Cl-cotransporter (Fragment).
 OS Eriocheir sinensis (Chinese mitten crab).
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
 OC Eumalacostraca; Decapoda; Pleocyemata; Brachyura;
 OC Eumalacostraca; Decapoda; Pleocyemata; Brachyura;
 OX NCBI_TaxID=95602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Gill;
 RA Wehrhach D., Towle D.W.;
 RT "Na+/H+-exchanger and Na+/K+/2Cl- cotransporter are expressed in
 RT gills of the euryhaline Chinese crab Eriocheir sinensis."
 RL Comp. Biochem. Physiol. 126:S158-S158(2000).
 DR EMBL; AF301160; AAG39938.1; -
 DR NON_TER 1
 FT NON_TER 219
 SQ SEQUENCE 219 AA; 24159 MW; 599442DA26FDD3F1 CRC64;

Query Match 64.8%; Score 35; DB 5; Length 219;
 Best Local Similarity 75.0%; Pred. No. 39;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 VVPGXGMDYS 11
 |||
 Db 107 VVPGXGMDYS 114

RESULT 14

O8XP8 PRELIMINARY; PRT; 253 AA.

AC O8XP8; ID O8XP8; PRELIMINARY; PRT; 253 AA.
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical protein CPE0057.
 GN CPE0057.
 OS Clostridium perfringens.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1502;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=13 / Type A;
 RX PubMed=11792842;
 RA Shimizu T., Ontani K., Hirakawa H., Ohehima K., Yamashita A.,
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic
 RT flesh-eater."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
 DR EMBL; AP003185; BAB79763.1; -
 DR InterPro; IPR002025; NAD_binding.
 DR InterPro; IPR000594; TnF_domain.
 DR Pfam; PF00899; TnF_1.
 DR Hypothetical protein; Complete proteome.
 SQ SEQUENCE 253 AA; 27714 MW; 814DF79D3BD7486 CRC64;

Query Match 64.8%; Score 35; DB 16; Length 253;
 Best Local Similarity 50.0%; Pred. No. 46;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 EEVPGXGMDY 10
 |||
 Db 108 EEVPGXGMDY 117

RESULT 15

O52367 PRELIMINARY; PRT; 298 AA.

AC O52367; ID O52367; PRELIMINARY; PRT; 298 AA.
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Aryl-alcohol dehydrogenase homolog (Fragment).
 GN XylB1.
 OS Rhizobium tropici.
 OG Rhizobium tropici.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiales; Rhizobium.
 OX NCBI_TaxID=398;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CFN299;
 RA Rosenblueth M., Hynes M.F., Martinez-Romero R.;
 RT Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- COFACTOR: ZINC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY.
 DR EMBL; AF036920; AAC04779.1; -
 DR InterPro; IPR00328; ADH_zinc.
 DR InterPro; IPR002085; Adh_zn family.
 DR InterPro; IPR00205; NAD_binding.

DR Pfam: PF00107; adh_zinc; 1.
DR PROSITE: PS00059; ADH_ZINC; 1.
KW Metal-binding; Oxidoreductase; Zinc; Plasmid.
FT NON_TER 298
SQ SEQUENCE 298 AA; 31092 MW; 49B2F817C33AE87 CRC64;

Query Match 64.8%; Score 35; DB 2; Length 298;
Best Local Similarity 50.0%; Pred No. 55;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 2 EVVFXGMDYS 11
|:|:|:|:|:
Db 250 EIRFGADFS 259

Search completed: December 22, 2003, 17:51:29
Job time : 28.2667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 16:41:00 ; Search time 32.4667 Seconds
(without alignments)
53.778 Million cell updates/sec

Title: US-09-909-164-9
Perfect score: 52
Sequence: 1 BEVVPXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
A.GeneSeq.190un03.*
1: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	50	96.2	11 23 ABB80521	Hepatitis C virus
2	50	96.2	11 23 ABB80522	Hepatitis C virus
3	50	96.2	11 23 ABB80525	Hepatitis C virus
4	50	96.2	11 23 ABB80526	Hepatitis C virus
5	50	96.2	11 23 ABB80529	Hepatitis C virus
6	50	96.2	11 23 ABB80559	Hepatitis C virus
7	50	96.2	11 23 ABB80563	Hepatitis C virus
8	50	96.2	11 23 ABB80564	Hepatitis C virus
9	50	96.2	11 23 ABB80565	Hepatitis C virus
			11 23 ABB80566	Hepatitis C virus

10	50	96.2	11 23 ABB80567	Hepatitis C virus
11	50	96.2	11 23 ABB80568	Hepatitis C virus
12	46	88.5	11 23 ABB80524	Hepatitis C virus
13	46	88.5	11 23 ABB80528	Hepatitis C virus
14	46	88.5	11 23 ABB80529	Hepatitis C virus
15	46	88.5	11 23 ABB80561	Hepatitis C virus
16	46	88.5	11 23 ABB80562	Hepatitis C virus
17	45	86.5	11 23 ABB80523	Hepatitis C virus
18	45	86.5	11 23 ABB80527	Hepatitis C virus
19	45	86.5	11 23 ABB80535	Hepatitis C virus
20	45	86.5	11 23 ABB80536	Hepatitis C virus
21	45	86.5	11 23 ABB80539	Hepatitis C virus
22	45	86.5	11 23 ABB80540	Hepatitis C virus
23	45	86.5	11 23 ABB80558	Hepatitis C virus
24	45	86.5	11 23 ABB80560	Hepatitis C virus
25	44	84.6	11 23 ABB80544	Hepatitis C virus
26	44	84.6	11 23 ABB80545	Hepatitis C virus
27	44	84.6	11 23 ABB80549	Hepatitis C virus
28	44	84.6	11 23 ABB80552	Hepatitis C virus
29	44	84.6	11 23 ABB80553	Hepatitis C virus
30	42	80.8	11 23 ABB80530	Hepatitis C virus
31	41	78.8	11 23 ABB80538	Hepatitis C virus
32	41	78.8	11 23 ABB80542	Hepatitis C virus
33	41	78.8	11 23 ABB80543	Hepatitis C virus
34	40	76.9	11 23 ABB80537	Hepatitis C virus
35	40	76.9	11 23 ABB80541	Hepatitis C virus
36	40	76.9	11 23 ABB80547	Hepatitis C virus
37	40	76.9	11 23 ABB80548	Hepatitis C virus
38	40	76.9	11 23 ABB80551	Hepatitis C virus
39	40	76.9	11 23 ABB80556	Hepatitis C virus
40	40	76.9	11 23 ABB80557	Hepatitis C virus
41	40	76.9	20 22 AAU76810	Hepatitis C virus
42	40	76.9	22 22 ABG03621	Hepatitis C virus
43	40	76.9	1022 22 ABG05826	Novel human diagno
44	40	76.9	1022 22 ABG08173	Novel human diagno
45	39	75.0	11 23 ABB80546	Hepatitis C virus

ALIGNMENTS

RESULT 1	
ABB80521	
ID	ABB80521 standard; peptide; 11 AA.
AC	ABB80521;
XX	
DT	08-OCT-2002 (first entry)
XX	
DE	Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.
XX	
KW	Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM	virucide.
XX	
OS	Synthetic.
XX	
FT	Key
FT	Modified-site 1 Location/Qualifiers
FT	Modified-site /note= "N-terminal acetyl"
FT	Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT	FT residue 7"
FT	Modified-site 11 /note= "C-terminal amide"
XX	
PN	WO200208251-A2.
XX	
PD	31-JAN-2002.
XX	
PF	19-JUL-2001; 2001WO-US23169.
XX	
PR	21-JUL-2000; 2000US-220101P.
XX	

PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
PS Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 1 BEVVPXGMSYS 11

RESULT 2
ABB80522
ID ABB80522 standard; peptide; 11 AA.
XX
AC ABB80522;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.
XX
KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "N-terminal acetyl" forming keto-amide linkage with
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PR 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C

PT virus protease -
XX
PS Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 1 BEVVPXGMSYS 11

RESULT 3
ABB80525
ID ABB80525 standard; peptide; 11 AA.
XX
AC ABB80525;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #5.
XX
KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "N-terminal acetyl" forming keto-amide linkage with
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PR 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
PS Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of

CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX

Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BBVVPXGMSYS 11
|||||
DB 1 BBVVPXGMSYS 11

RESULT 4
ABB80526
ID ABB80526 standard; peptide; 11 AA.

AC ABB80526;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #6.

KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
virocide.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory

XX activity useful for treating disorders associated with hepatitis C

XX virus protease -

XX Claim 17; Page 64; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virocidic activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BBVVPXGMSYS 11
|||||
DB 1 BBVVPXGMSYS 11

RESULT 5

ABB80559
ID ABB80559 standard; peptide; 11 AA.

AC ABB80559;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #39.

KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
virocide.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Modified-site 8 /note= "Oxymethionine"

FT Modified-site 11 /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C

XX virus protease -

XX Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virocidic activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BBVVPXGMSYS 11
|||||

Db 1 EEVVPXGMSYS 11

RESULT 6
ABB80563
ID ABB80563 standard; peptide; 11 AA.
XX
AC ABB80563;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #43.
XX
KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX
KM virucide.
XX
OS Synthetic.
XX
FH Key
FT Modified-site 1 Location/Qualifiers
FT /note= "N-terminal acetyl"
FT Modified-site 6
FT /note= "Valyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11
FT /note= "C-terminal amide"
XX
PN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
XX
XX virus protease -
XX
PS Claim 17; Page 65; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
XX
Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
DB 1 EEVVPXGMSYS 11

RESULT 7
ABB80564
ID ABB80564 standard; peptide; 11 AA.
XX
AC ABB80564;
XX
DT 08-OCT-2002 (first entry)

XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #44.
XX
KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX
KM virucide.
XX
OS Synthetic.
XX
FH Key
FT Modified-site 1 Location/Qualifiers
FT /note= "N-terminal acetyl"
FT Modified-site 6
FT /note= "Leucyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11
FT /note= "C-terminal amide"
XX
PN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PF Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
XX
XX virus protease -
XX
PS Claim 17; Page 65; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
XX
Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
DB 1 EEVVPXGMSYS 11

RESULT 8
ABB80565
ID ABB80565 standard; peptide; 11 AA.
XX
AC ABB80565;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #45.
XX
KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX
KM virucide.
XX
OS Synthetic.
XX
FH Key
FT Modified-site 1 Location/Qualifiers

```

FT      Modified-site      /note= "N-terminal acetyl"
FT      6
FT      /note= "Norleucyl carbonyl forming keto-amide linkage
FT      with residue 7"
FT      Modified-site      11
FT      /note= "C-terminal amide"
XX      WO200208251-A2.
XX      31-JAN-2002.
XX      19-JUL-2001; 2001WO-US23169.
XX      21-JUL-2000; 2000US-220101P.
XX      (CORV-) CORVAS INT INC.
XX      Lim-wilby M, Levy OE, Brunck TK;
XX      WPI, 2002-361643/39.
XX      Novel peptide compound having hepatitis C virus protease inhibitory
XX      activity useful for treating disorders associated with hepatitis C
XX      virus protease
XX      Claim 17; Page 65; 69pp; English.
XX      The sequence represents a peptide compound of the invention having
XX      hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX      invention are alpha-ketoamide peptide analogues. The peptides have
XX      virucide activity, and are useful for treating and in the manufacture of
XX      a medicament to treat disorders associated with HCV protease. A
XX      pharmaceutical composition comprising the peptide as an active ingredient
XX      is useful for treating disorders associated with hepatitis C virus.
XX      Sequence 11 AA;
XX      SQ
XX      Query Match      96.2%; Score 50; DB 23; Length 11;
XX      Best Local Similarity 100.0%; Pred. No. 0.0013;
XX      Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 EEVVPXGMSYS 11
DB      1 EEVVPXGMSYS 11
XX      RESULT 9
XX      ABB80566
XX      ID ABB80566 standard; peptide; 11 AA.
XX      AC ABB80566;
XX      DT 08-OCT-2002 (first entry)
XX      DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #46.
XX      KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX      virucide.
XX      OS Synthetic.
XX      Key Location/Qualifiers
XX      Modified-site 1 /note= "N-terminal acetyl"
XX      Modified-site 6 /note= "2-aminoisobutyryl carbonyl residue forming a
XX      Modified-site 11 keto-amide linkage with residue 7"
XX      /note= "C-terminal amide"
XX      WO200208251-A2.
XX      31-JAN-2002.

```

```

XX      19-JUL-2001; 2001WO-US23169.
XX      21-JUL-2000; 2000US-220101P.
XX      (CORV-) CORVAS INT INC.
XX      Lim-wilby M, Levy OE, Brunck TK;
XX      WPI, 2002-361643/39.
XX      Novel peptide compound having hepatitis C virus protease inhibitory
XX      activity useful for treating disorders associated with hepatitis C
XX      virus protease
XX      Claim 17; Page 65; 69pp; English.
XX      The sequence represents a peptide compound of the invention having
XX      hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX      invention are alpha-ketoamide peptide analogues. The peptides have
XX      virucide activity, and are useful for treating and in the manufacture of
XX      a medicament to treat disorders associated with HCV protease. A
XX      pharmaceutical composition comprising the peptide as an active ingredient
XX      is useful for treating disorders associated with hepatitis C virus.
XX      Sequence 11 AA;
XX      SQ
XX      Query Match      96.2%; Score 50; DB 23; Length 11;
XX      Best Local Similarity 100.0%; Pred. No. 0.0013;
XX      Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 EEVVPXGMSYS 11
DB      1 EEVVPXGMSYS 11
XX      RESULT 10
XX      ABB80567
XX      ID ABB80567 standard; peptide; 11 AA.
XX      AC ABB80567;
XX      DT 08-OCT-2002 (first entry)
XX      DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #47.
XX      KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX      virucide.
XX      OS Synthetic.
XX      Key Location/Qualifiers
XX      Modified-site 1 /note= "N-terminal acetyl"
XX      Modified-site 6 /note= "(e,s)allothreonyl carbonyl residue forming a
XX      Modified-site 11 keto-amide linkage with residue 7"
XX      /note= "C-terminal amide"
XX      WO200208251-A2.
XX      31-JAN-2002.
XX      19-JUL-2001; 2001WO-US23169.
XX      21-JUL-2000; 2000US-220101P.
XX      (CORV-) CORVAS INT INC.
XX      Lim-wilby M, Levy OE, Brunck TK;
XX      WPI, 2002-361643/39.

```

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX
PS Claim 17; Page 65; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EEVVPXGMSYS 11
DB 1 EEVVPXGMSYS 11
RESULT 11
ABB80568
ID ABB80568 standard; peptide; 11 AA.
XX
AC ABB80568;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #48.
XX
KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Alpha-propionyl-glycyl-carbonyl residue forming
FT a keto-amide linkage with residue 7"
FT Modified-site 11 /note= "C-terminal amide"
XX
PN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PP 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
PS WPI; 2002-361643/39.
XX
DR Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 65; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EEVVPXGMSYS 11
DB 1 EEVVPXGMSYS 11
RESULT 12
ABB80524
ID ABB80524 standard; peptide; 11 AA.
XX
AC ABB80524;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.
XX
KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KM virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
PN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PP 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
PS WPI; 2002-361643/39.
XX
DR Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 88.5%; Score 46; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.0088;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 11
|||
DB 1 EEVVPXGMDYS 11

RESULT 13

ABB80528
ID ABB80528 standard; peptide; 11 AA.

AC ABB80528;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #8.

KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.

OS Synthetic.

Key Location/Qualifiers

FT Modified-site 1

FT /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

RESULT 14
ABB80529
ID ABB80529 standard; peptide; 11 AA.

AC ABB80529;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.

KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.

OS Synthetic.

Key Location/Qualifiers

FT Modified-site 1

FT /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

Query Match 88.5%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0088;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 11
|||
DB 1 EEVVPXGMDYS 11

RESULT 15

ABB80561
ID ABB80561 standard; peptide; 11 AA.

AC ABB80561;

XX

```

DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #41.
XX
KM Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 8 /note= "Oxymethionine"
FT Modified-site 11 /note= "C-terminal amide"
FT
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease -
XX
XX Claim 17; Page 65; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a pharmaceutical to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;

```

Query Match 88.5%; Score 46; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.0088;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 EEVVPXGMSYS 11
   |||||
Db 1 EEVVPXGMDYS 11

```

Search completed: December 22, 2003, 17:41:00
 Job time : 32.4667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 16:37:03 / Search time 10.1333 Seconds
(without alignments)
45.930 Million cell updates/sec

Title: US-09-909-164-9

Perfect score: 52

Sequence: 1 BEVVPXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents: AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/6C.COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	65.4	947	4	US-09-228-986-73
2	65.4	1407	4	US-09-328-352-7885
3	63.5	45	2	US-08-637-759B-236
4	63.5	45	3	US-08-871-355A-236
5	63.5	45	4	US-09-201-945-236
6	63.5	65	6	5177197-51
7	63.5	410	6	5177197-1
8	63.5	1394	6	5177197-30
9	63.5	10	3	US-09-357-952-66
10	63.5	10	4	US-09-521-650-66
11	63.5	10	4	US-09-168-888-66
12	63.5	102	2	US-08-580-988A-23
13	63.5	152	2	US-08-460-694-4
14	63.5	152	3	US-08-460-744-4
15	63.5	153	3	US-07-667-711B-4
16	63.5	173	1	US-08-193-977-7
17	63.5	189	2	US-08-464-517-21
18	63.5	189	2	US-08-246-361A-21
19	63.5	189	3	US-08-463-772-21
20	63.5	189	5	PCT-US93-05000-21
21	63.5	236	2	US-08-464-517-22
22	63.5	236	2	US-08-246-361A-22
23	63.5	236	5	PCT-US93-05000-22
24	63.5	280	3	US-08-463-772-22
25	63.5	280	3	US-08-464-517-6
26	63.5	280	3	US-08-463-772-6
27	63.5	289	2	US-08-246-361A-4

28	32	61.5	289	5	PCT-US93-05000-4	Sequence 4, Appli
29	32	61.5	291	5	PCT-US93-05000-6	Sequence 6, Appli
30	32	61.5	292	2	US-08-464-517-23	Sequence 23, Appli
31	32	61.5	292	2	US-08-246-361A-6	Sequence 6, Appli
32	32	61.5	292	2	US-08-246-361A-23	Sequence 23, Appli
33	32	61.5	292	3	US-08-463-772-23	Sequence 23, Appli
34	32	61.5	292	1	US-07-947-140-8	Sequence 8, Appli
35	32	61.5	295	1	US-08-472-893A-8	Sequence 8, Appli
36	32	61.5	295	2	US-08-460-694-2	Sequence 2, Appli
37	32	61.5	295	2	US-08-464-517-19	Sequence 19, Appli
38	32	61.5	295	2	US-08-464-517-20	Sequence 20, Appli
39	32	61.5	295	2	US-08-246-361A-19	Sequence 19, Appli
40	32	61.5	295	2	US-08-246-361A-20	Sequence 20, Appli
41	32	61.5	295	3	US-08-463-772-19	Sequence 19, Appli
42	32	61.5	295	3	US-08-463-772-20	Sequence 20, Appli
43	32	61.5	295	3	US-08-460-744-2	Sequence 2, Appli
44	32	61.5	295	3	US-07-667-711B-2	Sequence 2, Appli
45	32	61.5	295	3	US-07-667-711B-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-228-986-73
Sequence 73, Application US/09228986
Patent No. 6359198
GENERAL INFORMATION:
APPLICANT: Strabala, Timothy
APPLICANT: Niemann, Niels
TITLE OF INVENTION: Compositions isolated from plant cells
TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling
FILE REFERENCE: 11000/1020
CURRENT APPLICATION NUMBER: US/09/228,986
CURRENT FILING DATE: 1999-01-12
NUMBER OF SEQ ID NOS: 130
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 73
LENGTH: 947
TYPE: PRT
ORGANISM: Pinus radiata

Query Match
Best Local Similarity 65.4%; Score 34; DB 4; Length 947;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
DB 686 VMPXGMSYS 694

RESULT 2
US-09-328-352-7885
Sequence 7885, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 7885
LENGTH: 1407
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-7885

Query Match
Best Local Similarity 65.4%; Score 34; DB 4; Length 1407;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 EVVFXGMSY 10
DB 596 EVPEGLSF 604

RESULT 3

US-08-637-759B-236
; Sequence 236, Application US/08637759B
; Patent No. 5876931
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,759B
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-637-759B-236

Query Match 63.5%; Score 33; DB 2; Length 45;
Best Local Similarity 60.0%; Pred. No. 6.6;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 EEVVPXGMSY 10
DB 1 EEISPLGMSY 10

RESULT 4
US-08-871-355A-236
; Sequence 236, Application US/08871355A
; Patent No. 6015669
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street

CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871,355A
; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101 CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-871-355A-236

Query Match 63.5%; Score 33; DB 3; Length 45;
Best Local Similarity 60.0%; Pred. No. 6.6;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 EEVVPXGMSY 10
DB 1 EEISPLGMSY 10

RESULT 5
US-09-201-945-236
; Sequence 236, Application US/09201945
; Patent No. 6342215
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/201,945
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/637,759
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:

NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPS 101
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8795
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 236:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-09-201-945-236

Query Match 63.5%; Score 33; DB 4; Length 45;
Best Local Similarity 60.0%; Pred. No. 6.6;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 10
: : : : :
Db 1 EESISPLGMSY 10

RESULT 6
5177197-51
PATENT NO. 5177197
APPLICANT: KANZAKI, TETSUO;OLOFSSON, ANDERS;MOREN, ANITA;
; WERNSTEDT, CHRISTER;HELLMAN, ULF;MIYAZONO, KOHEI;CLAESSON-WEISH,
; LENA;HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETA1-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO: 51:
; LENGTH: 65
5177197-51

Query Match 63.5%; Score 33; DB 6; Length 65;
Best Local Similarity 45.5%; Pred. No. 10;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 11
: : : : :
Db 52 KEICPGMGMYT 62

RESULT 7
5177197-1
PATENT NO. 5177197
APPLICANT: KANZAKI, TETSUO;OLOFSSON, ANDERS;MOREN, ANITA;
; WERNSTEDT, CHRISTER;HELLMAN, ULF;MIYAZONO, KOHEI;CLAESSON-WEISH,
; LENA;HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETA1-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO: 1:
; LENGTH: 410
5177197-1

Query Match 63.5%; Score 33; DB 6; Length 410;
Best Local Similarity 45.5%; Pred. No. 81;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 11
: : : : :
Db 399 KEICPGMGMYT 409

RESULT 8
5177197-30
PATENT NO. 5177197
APPLICANT: KANZAKI, TETSUO;OLOFSSON, ANDERS;MOREN, ANITA;
; WERNSTEDT, CHRISTER;HELLMAN, ULF;MIYAZONO, KOHEI;CLAESSON-WEISH,
; LENA;HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETA1-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO: 30:
; LENGTH: 1394
5177197-30

Query Match 63.5%; Score 33; DB 6; Length 1394;
Best Local Similarity 45.5%; Pred. No. 3.3e+02;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 11
: : : : :
Db 399 KEICPGMGMYT 409

RESULT 9
US-09-357-952-66
SEQUENCE 66, Application US/09357952
PATENT NO. 6248904
GENERAL INFORMATION:
APPLICANT: Zhang, Han-Zhong
APPLICANT: Cai, Sui Xiong
APPLICANT: Drewe, John A.
TITLE OF INVENTION: No. 6248904e1 Fluorescence Dyes and Their Applications for Who
TITLE OF INVENTION: Fluorescence Screening Assays for Caspases, Replicases, Proteas
TITLE OF INVENTION: Other Enzymes and the Use Thereof
FILE REFERENCE: 1735.0030001
CURRENT APPLICATION NUMBER: US/09/357,952
CURRENT FILING DATE: 1999-07-21
EARLIER APPLICATION NUMBER: US 60/093,642
EARLIER FILING DATE: 21-JUL-1998
NUMBER OF SEQ ID NOS: 139
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 66
LENGTH: 10
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-357-952-66

Query Match 61.5%; Score 32; DB 3; Length 10;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 10
: : : : :
Db 1 DDIVPCMSYS 10

RESULT 10
US-09-521-650-66
SEQUENCE 66, Application US/09521650
PATENT NO. 6335429
GENERAL INFORMATION:
APPLICANT: Weber, Eckard
APPLICANT: Cai, Sui Xiong
APPLICANT: Keana, John F.W.
APPLICANT: Drewe, John A.

APPLICANT: Zhang, Han-Zhong
TITLE OF INVENTION: No. 63542961 Fluorogenic or Fluorescent Reporter Molecules and
TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence
TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the
FILE REFERENCE: 1735.0290002
CURRENT APPLICATION NUMBER: US/09/521,650
EARLIER FILING DATE: 2000-03-08
EARLIER APPLICATION NUMBER: 09/168,888
EARLIER FILING DATE: 1998-10-09
EARLIER APPLICATION NUMBER: US 60/061,582
EARLIER FILING DATE: 1997-10-10
EARLIER APPLICATION NUMBER: US 09/033,661
EARLIER FILING DATE: 1998-03-03
NUMBER OF SEQ ID NOS: 142
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 66
LENGTH: 10
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-521-650-66

Query Match 61.5% Score 32; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
:::|:|:|
DB 1 DDIVPCMSY 10

RESULT 11
US-09-168-888-66
Sequence 66, Application US/09168888
Patent No. 6342611
GENERAL INFORMATION:
APPLICANT: Weber, Eckard
APPLICANT: Cai, Sui Xiong
APPLICANT: Keana, John F.W.
APPLICANT: Drewe, John A.
APPLICANT: Zhang, Han-Zhong
TITLE OF INVENTION: No. 63426111 Fluorogenic or Fluorescent Reporter Molecules and
TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence
TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the
FILE REFERENCE: 1735.0290002
CURRENT APPLICATION NUMBER: US/09/168,888
EARLIER FILING DATE: 1998-10-09
EARLIER APPLICATION NUMBER: US 60/061,582
EARLIER FILING DATE: 1997-10-10
EARLIER APPLICATION NUMBER: US 09/033,661
EARLIER FILING DATE: 1998-03-03
NUMBER OF SEQ ID NOS: 142
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 66
LENGTH: 10
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-168-888-66

Query Match 61.5% Score 32; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
:::|:|:|
DB 1 DDIVPCMSY 10

RESULT 12
US-08-580-988A-23
Sequence 23, Application US/08580988A
Patent No. 5856161
GENERAL INFORMATION:
APPLICANT: Aggarwal et al.
TITLE OF INVENTION: Tumor Necrosis Factor
TITLE OF INVENTION: Receptor-I-Associated Protein Kinase And Methods
TITLE OF INVENTION: For Its Use
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dr. Benjamin A. Adler
STREET: 8011 Candle Lane
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77071

COMPUTER READABLE FORM:
MEDIUM TYPE: 1.44 Mb floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word for Macintosh
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/580,988A
FILING DATE: January 3, 1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Benjamin Aaron Adler, Ph.D., J.D.
REGISTRATION NUMBER: 35,423
REFERENCE/DOCKET NUMBER: D5721CIP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-777-2321
TELEFAX: 713-777-6908
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 102 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: no
HYPOTHETICAL: no
ANTI-SENSE: no
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
US-08-580-988A-23

Query Match 61.5% Score 32; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
|||:|:|:|
DB 24 EEVFPPLAMNY 33

RESULT 13
US-08-460-694-4
Sequence 4, Application US/08460694
Patent No. 5858655
GENERAL INFORMATION:
APPLICANT: Arnold, Andrew
TITLE OF INVENTION: PRAD1 Cyclin and its cDNA
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington

```

; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,694
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McConathy, Evelyn H.
; REGISTRATION NUMBER: 35,279
; REFERENCE/DOCKET NUMBER: 0609.4070002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 152 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-460-694-4

Query Match      61.5%; Score 32; DB 2; Length 152;
Best Local Similarity 60.0%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy      1 EEVVPXGMSY 10
Db      20 EEVFPPLMNY 29

RESULT 14
US-08-460-744-4
; Sequence 4, Application US/08460744
; Patent No. 6107541
; GENERAL INFORMATION:
; APPLICANT: Arnold, Andrew
; TITLE OF INVENTION: PRAD1 Cyclin and its cDNA
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,744
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McConathy, Evelyn H.
; REGISTRATION NUMBER: 35,279
; REFERENCE/DOCKET NUMBER: 0609.4070005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2540
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 152 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; MOLECULE TYPE: not relevant
; 
```

```

; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-460-744-4

Query Match      61.5%; Score 32; DB 3; Length 152;
Best Local Similarity 60.0%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy      1 EEVVPXGMSY 10
Db      20 EEVFPPLMNY 29

RESULT 15
US-07-667-711B-4
; Sequence 4, Application US/07667711B
; Patent No. 6110700
; GENERAL INFORMATION:
; APPLICANT: ARNOLD, ANDREW
; TITLE OF INVENTION: Prad1 Cyclin and Its cDNA
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVE., NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/667,711B
; FILING DATE: 11-MAR-1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MCPHAIL, DONALD R.
; REGISTRATION NUMBER: 35,811
; REFERENCE/DOCKET NUMBER: 0609.4070000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2540
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 152 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; US-07-667-711B-4

Query Match      61.5%; Score 32; DB 3; Length 152;
Best Local Similarity 60.0%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy      1 EEVVPXGMSY 10
Db      20 EEVFPPLMNY 29

Search completed: December 22, 2003, 16:43:43
Job time : 12.1333 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 16:40:14 ; Search time 20.8667 Seconds
(without alignments)
98.451 Million cell updates/sec

Title: US-09-909-164-9
Perfect score: 52
Sequence: 1 EEVVPXGMSTYS 11

Scoring table:

BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Published Applications MA:
2: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	73.1	3472	12	US-10-029-120-4
2	38	73.1	3472	14	US-10-027-806-4
3	38	73.1	3472	14	US-10-034-623-4
4	38	73.1	3472	14	US-10-027-801-4
5	36	69.2	153	12	US-10-029-386-32486
6	35	67.3	236	12	US-10-029-386-32076
7	34	65.4	290	12	US-10-094-749-2565
8	34	65.4	947	15	US-10-101-464A-73
9	34	65.4	1163	12	US-10-122-067-4
10	34	65.4	1499	12	US-10-086-534-67
11	34	65.4	1499	12	US-10-122-067-2
12	33	63.5	99	15	US-10-106-698-7477
13	33	63.5	426	15	US-10-214-766-43
14	33	63.5	478	11	US-09-992-600A-108
15	33	63.5	478	11	US-09-924-340-108

16	33	63.5	478	11	US-09-746-783-184	Sequence 184, App
17	33	63.5	478	12	US-09-992-095B-108	Sequence 108, App
18	33	63.5	478	12	US-10-154-678-108	Sequence 108, App
19	33	63.5	478	12	US-09-999-570-108	Sequence 108, App
20	33	63.5	478	15	US-10-000-489-108	Sequence 108, App
21	33	63.5	478	15	US-10-000-986-108	Sequence 108, App
22	33	63.5	478	12	US-10-094-749-3010	Sequence 3010, App
23	33	63.5	653	11	US-09-820-843A-26	Sequence 26, App
24	33	63.5	1152	16	US-10-080-170-130	Sequence 130, App
25	33	63.5	1394	12	US-10-116-275-261	Sequence 261, App
26	32	61.5	10	10	US-09-947-387-66	Sequence 66, App
27	32	61.5	10	12	US-10-138-375-66	Sequence 66, App
28	32	61.5	36	12	US-10-199-820-241	Sequence 241, App
29	32	61.5	242	12	US-10-094-749-2076	Sequence 2076, App
30	32	61.5	254	9	US-09-778-927A-53	Sequence 53, App
31	32	61.5	276	12	US-10-116-275-114	Sequence 114, App
32	32	61.5	289	10	US-09-919-497-54	Sequence 54, App
33	32	61.5	289	14	US-10-024-066-2	Sequence 2, App
34	32	61.5	289	14	US-10-024-066-4	Sequence 4, App
35	32	61.5	292	12	US-10-116-275-275	Sequence 275, App
36	32	61.5	295	10	US-09-925-300-1061	Sequence 1061, App
37	32	61.5	529	9	US-09-923-304-4	Sequence 4, App
38	32	61.5	691	9	US-09-925-731-2	Sequence 2, App
39	32	61.5	691	15	US-10-101-921-4	Sequence 4, App
40	32	61.5	1377	9	US-09-815-242-10384	Sequence 10384, App
41	32	61.5	1377	12	US-10-287-274-467	Sequence 467, App
42	32	61.5	2799	14	US-10-151-736-4	Sequence 4, App
43	31	59.6	53	10	US-09-764-847-878	Sequence 878, App
44	31	59.6	53	15	US-10-092-154-878	Sequence 878, App
45	31	59.6	59	10	US-09-948-080-14	Sequence 14, App

ALIGNMENTS

RESULT 1
US-10-029-120-4
Sequence 4, Application US/10029120
Publication No. US20030175708A1
GENERAL INFORMATION:
APPLICANT: Swanson, Ronald V.
APPLICANT: Feldman, Robert A.
TITLE OR INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
FILE REFERENCE: DCCRP-002A
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/408,020
PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
NUMBER OF SEQ ID NOS: 123
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 3472
TYPE: PRT
ORGANISM: Cenarchaeum symbiosum
US-10-029-120-4

Query Match
Best Local Similarity 73.1%; Score 38; DB 12; Length 3472;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
DB 2294 EDVPRGRISPS 2304
QY 1 EEVVPXGMSTYS 11
US-10-027-806-4
Sequence 4, Application US/10027806
Publication No. US20020160476A1
GENERAL INFORMATION:
APPLICANT: Swanson, Ronald V.
APPLICANT: Feldman, Robert A.

```

; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCCP.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
; US-10-027-806-4

Query Match      73.1%; Score 38; DB 14; Length 3472;
Best Local Similarity 54.5%; Pred. NO. 1.6e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMSYS 11
Db      2294 EDVIRPGISFS 2304

RESULT 3
US-10-034-623-4
; Sequence 4, Application US/10034623
; Publication No. US20020198365A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCCP.002A
; CURRENT APPLICATION NUMBER: US/10/034,623
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
; US-10-034-623-4

Query Match      73.1%; Score 38; DB 14; Length 3472;
Best Local Similarity 54.5%; Pred. NO. 1.6e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMSYS 11
Db      2294 EDVIRPGISFS 2304

RESULT 4
US-10-027-801-4
; Sequence 4, Application US/10027801
; Publication No. US20030054364A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCCP.002A
; CURRENT APPLICATION NUMBER: US/10/027,801
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
```

```

; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
; US-10-027-801-4

Query Match      73.1%; Score 38; DB 15; Length 3472;
Best Local Similarity 54.5%; Pred. NO. 1.6e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY      1 EEVVPXGMSYS 11
Db      2294 EDVIRPGISFS 2304

RESULT 5
US-10-029-386-32486
; Sequence 32486, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32486
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF002994.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: SWISSPROT HIT: P43363, EVALU 5.00e-46
; US-10-029-386-32486

Query Match      69.2%; Score 36; DB 12; Length 153;
Best Local Similarity 70.0%; Pred. NO. 12;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 EEVVPXGMSY 10
Db      15 KEVVPFGHSY 24

RESULT 6
US-10-029-386-32076
; Sequence 32076, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32076
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

OTHER INFORMATION: MAP TO Z98050.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.3
OTHER INFORMATION: SWISSPROT HIT: P15822, EVALUATE 1.00e-125
US-10-029-386-32076

Query Match 67.3%; Score 35; DB 12; Length 236;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
DB 80 VVPAGLTYIS 88

RESULT 7

US-10-094-749-2565
Sequence 2565, Application US/10094749
Publication No. US20030219741A1

GENERAL INFORMATION:
APPLICANT: ISOGAI, TAKAO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: OTSUKI, TETSUJI
APPLICANT: WAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHICO
APPLICANT: YOSHIKAWA, TOSITOMU
APPLICANT: OTSUKA, MOTOTYUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
FILE REFERENCE: 084335/0160
CURRENT APPLICATION NUMBER: US/10/094,749
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 3381
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2565
LENGTH: 290
TYPE: PRT
ORGANISM: Homo sapiens
US-10-094-749-2565

Query Match 65.4%; Score 34; DB 12; Length 290;
Best Local Similarity 66.7%; Pred. No. 66;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGMS 9
DB 35 BEIVPMGIS 43

RESULT 8
US-10-101-464A-73
Sequence 73, Application US/10101464A
Publication No. US20030046728A1
GENERAL INFORMATION:

APPLICANT: Strabala, Timothy
APPLICANT: Nieuwenhuizen, Nicolaas
APPLICANT: Higgins, Colleen M.
TITLE OF INVENTION: Compositions Isolated from Plant Cells
TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signaling
FILE REFERENCE: 11000.1020C2
CURRENT APPLICATION NUMBER: US/10/101,464A
CURRENT FILING DATE: 2002-03-18
PRIOR APPLICATION NUMBER: 09/704,302
PRIOR FILING DATE: 2000-11-01
PRIOR APPLICATION NUMBER: 09/228,986
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/162,866
PRIOR FILING DATE: 1999-11-01
PRIOR APPLICATION NUMBER: PCT/US00/00724
PRIOR FILING DATE: 2000-01-11
NUMBER OF SEQ ID NOS: 989
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 73
LENGTH: 947
TYPE: PRT
ORGANISM: Pinus radiata
US-10-101-464A-73

Query Match 65.4%; Score 34; DB 15; Length 947;
Best Local Similarity 66.7%; Pred. No. 2,5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
DB 686 VMPGSGISYS 694

RESULT 9

US-10-122-067-4
Sequence 4, Application US/10122067
Publication No. US20030165883A1
GENERAL INFORMATION:
APPLICANT: ROY, A.J. Curtis
TITLE OF INVENTION: 27091, A PHOSPHOLIPID TRANSPORTING
TITLE OF INVENTION: ATPASE MOLECULE AND USES THEREFOR
FILE REFERENCE: MP101-058P1R
CURRENT APPLICATION NUMBER: US/10/122,067
CURRENT FILING DATE: 2002-04-12
PRIOR APPLICATION NUMBER: 60/283434
PRIOR FILING DATE: 2001-04-12
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 1163
TYPE: PRT
ORGANISM: Homo sapiens
US-10-122-067-4

Query Match 65.4%; Score 34; DB 12; Length 1163;
Best Local Similarity 72.7%; Pred. No. 3.1e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 133 BEVPRGSGSYS 143

RESULT 10
US-10-096-534-67
Sequence 67, Application US/10096534
Publication No. US20030166887A1
GENERAL INFORMATION:
APPLICANT: The Brigham and Women's Hospital, Inc.
APPLICANT: Yates, Karen
APPLICANT: Mizuno, Shuichi
APPLICANT: Glowacki, Julie
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF SKELETAL DEGENERATION CONDITIONS

```
; FILE REFERENCE: B0801/7244/XA/ERP
; CURRENT APPLICATION NUMBER: US/10/096, 534
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: US 60/274, 980
; PRIOR FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 67
; LENGTH: 1499
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-096-534-67
```

```
Query Match      65.4%; Score 34; DB 12; Length 1499;
Best Local Similarity 72.7%; Pred. No. 4.2e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 EEVVPXGMSYS 11
      |||||
Db      469 EEVVPXGMSYS 479
```

```
RESULT 11
US-10-122-067-2
; Sequence 2, Application US/10122067
; Publication No. US20030165883a1
; GENERAL INFORMATION:
; APPLICANT: RORY A.J. CURTIS
; TITLE OF INVENTION: 27091, A PHOSPHOLIPID TRANSPORTING
; FILE REFERENCE: MP101-058P1R
; CURRENT APPLICATION NUMBER: US/10/122, 067
; CURRENT FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/283434
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1499
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-122-067-2
```

```
Query Match      65.4%; Score 34; DB 12; Length 1499;
Best Local Similarity 72.7%; Pred. No. 4.2e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 EEVVPXGMSYS 11
      |||||
Db      469 EEVVPXGMSYS 479
```

```
RESULT 12
US-10-106-698-7477
; Sequence 7477, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106, 698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157, 137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163, 280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 7477
; LENGTH: 99
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (3)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (13)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (14)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (39)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (40)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (57)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (58)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (63)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (65)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (82)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (93)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-7477
```

```
Query Match      63.5%; Score 33; DB 15; Length 99;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      3 VVPXGMSYS 11
      |||||
Db      73 LVVPXGMSYS 81
```

```
RESULT 13
US-10-214-766-43
; Sequence 43, Application US/10214766
; Publication No. US20030084473A1
; GENERAL INFORMATION:
; APPLICANT: Gocal, Greg
; TITLE OF INVENTION: NON-TRANSGENIC HERBICIDE RESISTANT PLANTS
; FILE REFERENCE: CA1138
; CURRENT APPLICATION NUMBER: US/10/214, 766
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311, 734
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 43
; LENGTH: 426
; TYPE: PRT
; ORGANISM: Vibrio cholerae
US-10-214-766-43
```

```
Query Match      63.5%; Score 33; DB 15; Length 426;
Best Local Similarity 60.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 EEVVPXGMSYS 10
      |||||
Db      223 EFVVPXGMSYS 232
```

RESULT 14

US-09-992-600A-108
; Sequence 108, Application US/09992600A
; Publication No. US20030027161A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US4.DIV
; CURRENT APPLICATION NUMBER: US/09/992,600A
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: JPatent
; SEQ ID NO: 108
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-992-600A-108

Query Match

Best Local Similarity 63.5%; Score 33; DB 11; Length 478;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy

2 EVVPXGMSYS 11
|||
Db 239 EVAPAGASYN 248

RESULT 15

US-09-924-340-108
; Sequence 108, Application US/09924340
; Publication No. US20030027248A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US2.REG
; CURRENT APPLICATION NUMBER: US/09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO: 108
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-924-340-108

Query Match

Best Local Similarity 63.5%; Score 33; DB 11; Length 478;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 EVVPXGMSYS 11
|||

Db 239 EVAPAGASYN 248

Search completed: December 22, 2003, 17:32:42
Job time : 20.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using BW model

Run on: December 22, 2003, 17:24:36 ; Search time 9.06667 Seconds
(without alignments)

116.675 Million cell updates/sec

Title: US-09-909-164-9

Perfect score: 52

Sequence: 1 EEVVPXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	73.1	3472	2	T31308
2	37	71.2	840	2	T39116
3	37	71.2	877	2	T40413
4	36	69.2	102	2	A42452
5	36	69.2	1498	2	B97355
6	35	67.3	225	2	S57810
7	35	67.3	425	2	T24111
8	35	67.3	670	2	S22293
9	35	67.3	749	2	H82691
10	35	67.3	2717	2	A34203
11	34	65.4	156	2	S54619
12	34	65.4	252	2	HC9491
13	34	65.4	544	2	C82900
14	33	63.5	94	2	I40758
15	33	63.5	116	2	B90544
16	33	63.5	165	2	D69493
17	33	63.5	253	2	C81374
18	33	63.5	259	2	T34536
19	33	63.5	284	2	S75817
20	33	63.5	298	2	T47670
21	33	63.5	368	2	F72281
22	33	63.5	426	2	D82163
23	33	63.5	466	2	T34653
24	33	63.5	653	2	D82352
25	33	63.5	890	2	A30481
26	33	63.5	1028	2	AF3286
27	33	63.5	1152	2	D87046
28	33	63.5	1394	2	A35626
29	33	63.5	1401	2	G82336

30	33	63.5	1548	2	T04456	hypothetical prote
31	33	63.5	1712	2	A38261	making protein pr
32	32	61.5	84	2	B97333	hypothetical prote
33	32	61.5	175	2	P00616	transport protein
34	32	61.5	223	2	T01457	rho protein GDP-di
35	32	61.5	279	2	B72481	hypothetical prote
36	32	61.5	288	2	JC4011	cyclin D2 - rat
37	32	61.5	288	2	I58372	cyclin D2 - rat
38	32	61.5	289	2	A41984	cyclin D2 - mouse
39	32	61.5	289	2	A42822	cyclin D2 - human
40	32	61.5	291	2	S57922	cyclin D1 - Africa
41	32	61.5	291	2	S57925	cyclin D1 - Africa
42	32	61.5	291	2	JC4579	cyclin D1 - chicken
43	32	61.5	291	2	S62730	cyclin D1 - zebra
44	32	61.5	292	2	B42822	cyclin D3 - human
45	32	61.5	295	2	A38977	cyclin D1 - human

ALIGNMENTS

RESULT 1

T31308
hypothetical 367K protein - Cenarchaeum symbiosum
C:Species: Cenarchaeum symbiosum
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C:Accession: T31308
R:Schleper, C.; Delong, E.F.; Preston, C.M.; Feldman, R.A.; Wu, K.Y.; Swanson, R.V.
J. Bacteriol. 180, 5003-5009, 1998
A:Title: Genomic analysis reveals chromosomal variation in natural populations of the
A:Reference number: Z20994; MUID:98422450; PMID:9748430
A:Accession: T31308
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3472 <SCH>
A:Cross-references: EMBL:AF083072; NID:93599393; PID:93599394; PIDN:AAC62699.1
C:Superfamily: Cenarchaeum symbiosum hypothetical 367K protein

Query Match 73.1%; Score 38; DB 2; Length 3472;
Best Local Similarity 54.5%; Pred. No. 59;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
Db 2294 EDVIPRGISFS 2304

RESULT 2

T39116
probable sulfate permease - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T39116
R:Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, November 1999
A:Reference number: Z21829
A:Accession: T39116
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-840 <HUN>
A:Cross-references: EMBL:AI132779; PIDN:CA860015.1; GSPDB:GN00066; SPDB:SPAC869.05C
A:Experimental source: strain 972h-; cosmid c869
A:Gene: SPDB:SPAC869.05C
A:Map position: 1

Query Match 71.2%; Score 37; DB 2; Length 840;
Best Local Similarity 77.8%; Pred. No. 21;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMSYS 11
Db 135 VVPXGMSYA 143

RESULT 3
T40413
sulfate permease - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40413
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Jimenez Martinez, J.
submitted to the EMBL Data Library, August 1998
A:Reference number: Z1926
A:Accession: T40413
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-877 <LVN>
A:Cross-references: EMBL:AL031261; PIDN:CAA20298.1; GSPDB:GN00067; SPDB:SPBC3H7.02
A:Experimental source: strain 972h-; cosmid C3H7
C:Genetics:
A:Gene: SPDB:SPBC3H7.02
A:Map position: 2

Query Match
Best Local Similarity 71.2%; Score 37; DB 2; Length 877;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 EVVPXGMSYS 11
Db 148 VVPOGMSYA 156

RESULT 4
A42452
VI protein - tobacco yellow dwarf virus (strain Australia)
C:Species: tobacco yellow dwarf virus
C>Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C:Accession: A42452
R:Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
Virology 187, 633-642, 1992
A:Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow
A:Reference number: A42452; MUID:92188538; PMID:1546458
A:Accession: A42452
A:Molecule type: DNA
A:Residues: 1-102 <MOR>
A:Cross-references: GB:M01103; NID:G335283; PIDN:AAA47947.1; PID:G335284

Query Match
Best Local Similarity 69.2%; Score 36; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 3.5;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 2 EVVPXGMSYS 11
Db 7 QVVPXGMSYS 16

RESULT 5
B97355
DNA segregation ATPase, FtsK/SpoIIIE family, YUKA B. subtilis ortholog (imported) - Clos
C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: B97355
R:Rolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:11359325
A:Accession: B97355
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1498 <KOR>
A:Cross-references: GB:AE001437; PIDN:AAK81629.1; PID:G15026814; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC3709

Query Match
Best Local Similarity 69.2%; Score 36; DB 2; Length 1498;
Best Local Similarity 60.0%; Pred. No. 63;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 BEVVPXGMSY 10
Db 1276 EQKIPXGMSY 1285

RESULT 6
S57810
hypothetical protein precursor (clone TPPI) - tomato
C:Species: Lycopersicon esculentum (tomato)
C>Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
C:Accession: S57810
R:Milligan, S.B.; Gaesser, C.S.
Plant Mol. Biol. 28, 691-711, 1995
A:Title: Nature and regulation of pisp1l-expressed genes in tomato.
A:Reference number: S57808; MUID:95375233; PMID:7647301
A:Accession: S57810
A:Status: preliminary; nucleic acid sequence not shown
A:Reference number: S57808; MUID:95375233; PMID:7647301
A:Molecule type: mRNA
A:Residues: 1-225 <ML>
A:Cross-references: EMBL:U00592; NID:G924625; PIDN:AA80497.1; PID:G924626
C:Superfamily: plant Kunitz-type proteinase inhibitor

Query Match
Best Local Similarity 67.3%; Score 35; DB 2; Length 225;
Best Local Similarity 54.5%; Pred. No. 13;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 BEVVPXGMSYS 11
Db 32 DEVVPXGMSYA 42

RESULT 7
T24111
hypothetical protein R10D12.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24111
R:Percy, C.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19842
A:Accession: T24111
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-425 <ML>
A:Cross-references: EMBL:Z81109; PIDN:CAE03241.1; GSPDB:GN00023; CESP:R10D12.10
A:Experimental source: clone R10D12
C:Genetics:
A:Gene: CESP:R10D12.10
A:Map position: 5
A:Introns: 23/3; 56/3; 113/3; 257/2

Query Match
Best Local Similarity 67.3%; Score 35; DB 2; Length 425;
Best Local Similarity 50.0%; Pred. No. 26;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 BEVVPXGMSY 10
Db 335 EQVPGGLQY 344

RESULT 8
S22293
zinc finger protein AT-BP2 - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C>Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 20-Sep-1999
C:Accession: S22293; I78656
R:Mitchellmore, C.; Traboni, C.; Cortese, R.
Nucleic Acids Res. 19, 141-147, 1991

A>Title: Isolation of two cDNAs encoding zinc finger proteins which bind to the alpha 1-
A:Reference number: 158280; MUID:91187610; PMID:1901405
A:Accession: S22293
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-670 <MTR>
A:Cross-references: EMBL:X54250; NID:957519; PIDN:CAA38151.1; PID:957520
A>Note: the authors did not translate the codon for residue 1
C:Superfamily: HIV-SP2 enhancer-binding protein
C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 670;
Best Local Similarity 66.7%; Pred. No. 43;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 3 VVPXGMSYS 11
Db 376 VVPAGLTYS 384

RESULT 9
H82691
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: H82691
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen-
Nature 406, 151-157, 2000
A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A>Note: for a complete list of authors see reference number A59328 below
A:Accession: H82691
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-749 <SIM>
A:Cross-references: GB:AE003967; GB:AE003849; NID:91063347; PIDN:AAF84162.1; GSPDB:GN001
A:Experimental source: strain 945c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A-
Brienes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carver, H-
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Prohm
J.D.; Junqueira, M.L.; Kempner, E.L.; Kitajima, J.P.; Kielegier, J.E.; Kurume, E.E.; Laigt
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, B-
A:Authors: Martins, B.M.F.; Matukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XPI353
C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain A; phage T4 DNA topoisomerase

Query Match 67.3%; Score 35; DB 2; Length 749;
Best Local Similarity 77.8%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 EVVXGMSYS 10
Db 526 EVDPGMSYS 534

A>Title: A DNA-binding protein containing two widely separated zinc finger motifs that
A:Reference number: A34203; MUID:90165514; PMID:2106471
A:Accession: A34203
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2717 <FAN>
A:Cross-references: EMBL:X51435; NID:938017; PIDN:CAA5798.1; PID:938018
R:Balwin Jr., A.S.; Lecalat, K.P.; Singh, H.; Sharp, P.A.
Mol. Cell. Biol. 10, 1406-1414, 1990
A>Title: A large protein containing zinc finger domains binds to related sequence ele-
A:Reference number: A34779; MUID:90205817; PMID:2108316
A:Accession: A34779
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 801-1072, 'N', 1074-1168, 'K', 1170-1225, 'V', 1227-1434, 'N', 1436-1607, 'I', 1609.
A:Cross-references: GB:M32019
C:Superfamily: HIV-SP2 enhancer-binding protein
C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 2717;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 3 VVPXGMSYS 11
Db 2405 VVPAGLTYS 2413

RESULT 11
S54619
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C:Accession: S54619; S56879
R:de Haan, M.; Maarse, A.C.; Grivell, L.A.
submitted to the EMBL Data Library, May 1995
A:Reference number: S54619
A:Accession: S54619
A:Molecule type: DNA
A:Residues: 1-156 <DEH>
A:Cross-references: EMBL:X67331; NID:91041652; PIDN:CAA60762.1; PID:9829123
R:de Haan, M.; Grivell, L.A.; Maarse, A.C.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S56879
A:Accession: S56879
A:Molecule type: DNA
A:Residues: 1-156 <DEH>
A:Cross-references: EMBL:Z74920; NID:91420109; PIDN:CAA99201.1; PID:91420111; MIPS:YOF
A:Experimental source: strain S288C
C:Genetics:
A:Cross-references: SGD:S0005539
A:Map position: 15R
C:Superfamily: hypothetical protein YOR013w

Query Match 65.4%; Score 34; DB 2; Length 156;
Best Local Similarity 66.7%; Pred. No. 14;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 2 EVVXGMSYS 10
Db 50 EVVPLGMDY 58

RESULT 12
H69491
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 26-Aug-1999
C:Accession: H69491
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dode
Fleischmann, R.D.; Overbeek, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirschner, E.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997
 A:Authors: Uteerback, T.; Cotton, M.D.; Spriggs, T.; Artich, P.; Kaine, B.P.; Sykes, S.; Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaea
 A:Reference number: A69250; MUID:98049343; PMID:9389475
 A:Accession: H69491
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-252 <K8>
 A:Cross-references: GB:AE000970; GB:AE000782; NID:G2689293; PIDN:ABE89318.1; PID:G264860
 C:Superfamily: cell division inhibitor mind

Query Match 65.4%; Score 34; DB 2; Length 252;
 Best Local Similarity 75.0%; Pred. No. 24;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVFXGMS 9
 DB 81 EVVFXGMS 88

RESULT 13

probable ABC substrate-binding protein, iron UUS59 [imported] - Ureaplasma urealyticum
 C:Species: Ureaplasma urealyticum
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: C82900
 R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Caswell, G.H.
 A:Submitted to GenBank, February 2000
 A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mit
 A:Reference number: A82870
 A:Accession: C82900
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-544 <G1A>
 A:Cross-references: GB:AE002133; GB:AF222894; NID:G6899339; PIDN:AAF30768.1; GSPDB:GN001
 A:Experimental source: serovar 3; biovar 1
 C:Genetics:
 A:Gene: ABCbp-5; UUS59
 A:Genetic code: SGC3

Query Match 65.4%; Score 34; DB 2; Length 544;
 Best Local Similarity 70.0%; Pred. No. 55;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMSY 10
 DB 135 EEVVPXGMSY 144

RESULT 14

hypothetical protein 1 - Campylobacter jejuni (fragment)
 C:Species: Campylobacter jejuni
 C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
 C:Accession: I40758; S47317
 R:Hani, E.K.; Chan, V.L.
 J. Bacteriol. 177, 2396-2402, 1995
 A:Title: Expression and characterization of Campylobacter jejuni benzoylglutamine amidohy
 A:Reference number: I40758; MUID:95247673; PMID:7730270
 A:Accession: I40758
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-94 <RES>
 A:Cross-references: EMBL:Z36940; NID:G535805; PIDN:CAA85392.1; PID:G535806

Query Match 63.5%; Score 33; DB 2; Length 94;
 Best Local Similarity 55.6%; Pred. No. 14;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVFXGMSY 10
 DB 26 EVVFXGMSY 34

RESULT 15

50S ribosomal protein L20 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
 E90544
 C:Species: Mycoplasma pulmonis
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
 C:Accession: E90544
 R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallison, F.; Moszer, I.
 Nucleic Acids Res. 29, 2145-2153, 2001
 A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pu
 A:Reference number: A9512; MUID:21267165; PMID:11353084
 A:Accession: E90544
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-116 <KUR>
 A:Cross-references: GB:AL445566; PID:G14089674; PIDN:CAC13434.1; GSPDB:GN00153
 A:Experimental source: strain UAB CTIP
 C:Genetics:
 A:Gene: MPRU_2610
 A:Genetic code: SGC3
 C:Superfamily: Escherichia coli ribosomal protein L20

Query Match 63.5%; Score 33; DB 2; Length 116;
 Best Local Similarity 77.8%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
 DB 68 VVPXGMSYS 76

Search completed: December 22, 2003, 17:44:57
 Job time : 9.06667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: December 22, 2003, 16:43:51 / Search time 4.6 Seconds

(without alignments)
112.455 Million cell updates/sec

Title: US-09-909-164-9

Sequence: 1 EEVVPXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	73.1	1058	CARB_FUSNN	Q8R86 fuscobacteri
2	37	71.2	877	SULH_SCHPO	O74377 schizosacch
3	36	69.2	102	Y1IK_TYDVA	P31619 tobacco yel
4	36	69.2	1498	Y1A9_CLOAB	O04351 clostridium
5	35	67.3	2717	ZEPI_HUMAN	P15822 homo sapien
6	34	65.4	788	CY14_NEUCR	P23622 neurospora
7	34	65.4	1499	AL0C_HUMAN	O60312 homo sapien
8	33	63.5	116	RL20_MYCPU	O98470 mycoplasma
9	33	63.5	165	YJ49_ARCFU	O28330 archaeoglob
10	33	63.5	253	Y990_CAMJE	P45489 campylobact
11	33	63.5	280	CTX3_MOUSE	Q9K187 mus musculu
12	33	63.5	426	AROA_VIBCH	Q9K180 vibrio chol
13	33	63.5	466	CC37_SCHPO	O94740 schizosacch
14	33	63.5	478	GSR2_HUMAN	O9N2M5 homo sapien
15	33	63.5	890	BCN5_CLOPE	P08636 clostridium
16	33	63.5	1389	L7BS_MOUSE	O08318 mus musculu
17	33	63.5	1394	L7BS_HUMAN	P22054 homo sapien
18	33	63.5	1401	RPOC_VIBCH	Q9K293 vibrio chol
19	33	63.5	1595	L7BL_HUMAN	O14766 homo sapien
20	33	63.5	1712	L7BI_RAT	O00918 rattus norv
21	33	63.5	1713	L7BL_MOUSE	O08319 mus musculu
22	32	61.5	288	CGD2_RAT	O04827 rattus norv
23	32	61.5	289	CGD2_HUMAN	P30279 homo sapien
24	32	61.5	289	CGD2_MOUSE	P30280 mus musculu
25	32	61.5	291	CGD1_BRARE	O90459 brachydanio
26	32	61.5	291	CGD1_XENLA	P50755 xenopus lae
27	32	61.5	291	CGD2_CHICK	P49706 gallus gall
28	32	61.5	291	CGD2_XENLA	P53782 xenopus lae
29	32	61.5	292	CGD1_CHICK	P55169 gallus gall
30	32	61.5	292	CGD3_HUMAN	P30281 homo sapien
31	32	61.5	295	CGD1_HUMAN	P24385 homo sapien
32	32	61.5	295	CGD1_MOUSE	P25322 mus musculu
33	32	61.5	295	CGD1_RAT	P39948 rattus norv

34	32	61.5	427	1	TOLE_HAEIN	P44677 haemophilus
35	32	61.5	529	1	ENP3_HUMAN	O75355 homo sapien
36	32	61.5	691	1	S216_HUMAN	O95616 homo sapien
37	32	61.5	719	1	GSP_CRIEA	P90518 crithidia f
38	32	61.5	726	1	PRTF_HSV6U	P52384 human herpe
39	32	61.5	759	1	SC11_YEAST	P32784 saccharomyc
40	32	61.5	920	1	EBD_RAT	P28931 tomato aspe
41	32	61.5	993	1	VIA_TAV	P16916 escherichia
42	32	61.5	1377	1	RHSA_ECOLI	P16916 escherichia
43	32	61.5	1397	1	RHSC_ECOLI	P16917 escherichia
44	32	61.5	1411	1	RHSE_ECOLI	O95071 homo sapien
45	32	61.5	2799	1	EDD_HUMAN	

ALIGNMENTS

RESULT 1
AC CARB_FUSNN STANDARD; PRT; 1058 AA.
ID CARB_FUSNN
AC Q8R86;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthase ammonia chain).
OS CARB OR FNO422.
OC Fuscobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A., Bhatnagar A., Bartman A., Gardner W., Grechkin G., Zhu L., Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A., Larsen N., D'Souza M., Malinas T., Busch G., Haselkorn R., Ronselein M., Kyrpides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.
CC -1- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -1- PATHWAY: Arginine biosynthesis.
CC -1- PATHWAY: Pyrimidine biosynthesis; first step.
CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).
CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC EMBL: A010554; A0194625.1; ALT_INT.
DR HAMAP; MF_01210; -, 1.
DR InterPro; IPR006275; CarL_glu.
DR InterPro; IPR005483; Cpsae_L.
DR InterPro; IPR005479; Cpsae_L_D2.
DR InterPro; IPR005480; Cpsae_L_D3.
DR InterPro; IPR005481; Cpsae_L_N.
DR InterPro; IPR004362; MGS_Ilike.
DR Pfam; PF00289; Cpsase_L_chain; 2.
DR Pfam; PF02786; Cpsase_L_D2; 2.
DR Pfam; PF02787; Cpsase_L_D3; 1.

```

DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPSASE.
DR TIGRfam; TIGR01369; CPSase_1; 1.
DR PROSITE; PS00867; CPSASE_1; 2.
DR ARGININE BIOSYNTHESIS; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1058
FT NP_BIND 153 210
FT NP_BIND 302 352
FT METAL 284 284
FT METAL 298 298
FT METAL 300 300
FT METAL 820 820
FT METAL 832 832
SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1E39F CRC64;

Query Match 73.1%; Score 38; DB 1; Length 1058;
Best Local Similarity 60.0%; Pred. No. 6.3;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVFXGMSYS 11
DB 190 EIVPNGIAYS 199

RESULT 2
SULH_SCHPO STANDARD; PRT; 877 AA.
ID SULH_SCHPO
AC 074377;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable sulfate permease C3H7.02.
GN SPBC3H7.02.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetaceae; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
ON NCBI_TaxID=4896;
RX SEQUENCE FROM N.A.
RP STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkert G., Aert R., Robben J., Gymnopoulos B.,
RA Willems I., Vancraeynest E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Mambuti R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hunt S.M.,
RA Lucas R., Rochet M., Galliard C., Tallard V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

```

```

RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002)
CC -1- FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE INTO THE CELL (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -1- SIMILARITY: BELONGS TO THE SLC26A FAMILY OF TRANSPORTERS.
CC -1- SIMILARITY: Contains 1 STAS domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AL031261; CAA20298.1; -.
CC F1R; T40413; T40413.
CC GeneDB; Spombe; SPBC3H7.02; -.
CC InterPro; IPR002645; STAS.
CC InterPro; IPR001902; Sulph_transpt.
CC Pfam; PF01740; STAS; 1.
CC Pfam; PF00916; Sulfate_transp; 1.
CC TIGRfam; TIGR00815; sulp; 1.
CC PROSITE; PS01130; SLC26A; 1.
CC PROSITE; PS0801; STAS; 1.
CC
CC KW Transport; Transmembrane.
FT DOMAIN 133 153
FT TRANSMEM 161 181
FT TRANSMEM 181 181
FT TRANSMEM 186 206
FT TRANSMEM 221 241
FT TRANSMEM 243 263
FT TRANSMEM 292 312
FT TRANSMEM 329 349
FT TRANSMEM 384 404
FT TRANSMEM 424 444
FT TRANSMEM 461 481
FT TRANSMEM 484 504
FT TRANSMEM 518 538
FT TRANSMEM 543 563
FT TRANSMEM 594 747
FT DOMAIN 747 747
SQ SEQUENCE 877 AA; 96373 MW; 56995A8493371E43 CRC64;

Query Match 71.2%; Score 37; DB 1; Length 877;
Best Local Similarity 77.8%; Pred. No. 8.4;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVFXGMSYS 11
DB 148 VVFXGMSYS 156

RESULT 3
Y1LK_TYDVA STANDARD; PRT; 102 AA.
ID Y1LK_TYDVA
AC P31619;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Hypothetical 11.2 kDa protein.
GN Y1.
OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
OC VIRUSES; ssDNA viruses; Geminiviridae; Mastrevirus.
ON NCBI_TaxID=31599;
RX SEQUENCE FROM N.A.
RC MEDLINE=92188538; PubMed=1546458;
RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;
RA "The nucleotide sequence of the infectious cloned DNA component of
RA tobacco yellow dwarf virus reveals features of geminiviruses
RA infecting monocotyledonous plants.";
RL Virology 187:633-642(1992).
CC

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: M81103: AAA47947.1; -.
 DR PIR: A42452: A42452.
 DR InterPro: IPR002621: Gemini_mov.
 DR Pfam: PF01708: Gemini_mov.1.
 KW Hypothetical protein.
 SQ SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF55B67 CRC64;

Query Match 69.2%; Score 36; DB 1; Length 102;
 Best Local Similarity 60.0%; Pred. No. 1.4;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVVPXGMSYS 11
 Db 7 QVVPSTGINS 16

RESULT 4
 ID Y1A9_CLOAB STANDARD; PRT; 1498 AA.
 AC Q04351;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hypothetical protein CAC3709.
 GN CAC3709
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21359325; PubMed=1146286;
 RA Noelling J., Berton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.,
 RT Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum".
 RL J. Bacteriol. 183:4823-4838 (2001).
 RN [2]
 RP SEQUENCE OF 1-108 FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=93273706; PubMed=8501044;
 RA Sauer U., Duerte P.;
 RT "Sequence and molecular characterization of a DNA region encoding a
 RT small heat shock protein of Clostridium acetobutylicum".
 RL J. Bacteriol. 175:3394-3400 (1993).
 CC -1- SIMILARITY: Contains 2 Pfam/SpoIIIE domains.
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
 CC in positions 76 and 106.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: A8007866; AAX81629.1; -.
 DR EMBL: X65276; CAA46379.1; ALT_FRAME.
 DR PIR: B97355; B97355.
 DR InterPro: IPR002543; FtsK_SpoIIIE.
 DR Pfam: PF01580; FtsK_SpoIIIE; 2.

DR PROSITE; PS50901; FtsK_SpoIIIE; 2.
 KW Hypothetical protein; ATP-binding; Complete proteome; Repeat.
 FT DOMAIN 655 857 FtsK/SpoIIIE 1.
 FT NP_BIND 675 682 ATP (POTENTIAL).
 FT DOMAIN 1001 1188 FtsK/SpoIIIE 2.
 SQ SEQUENCE 1498 AA; 168968 MW; PF42037A35A9649 CRC64;

Query Match 69.2%; Score 36; DB 1; Length 1498;
 Best Local Similarity 60.0%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKVVPXGMSY 10
 Db 1276 EQKIPWGMGY 1285

RESULT 5
 ID ZEP1_HUMAN STANDARD; PRT; 2717 AA.
 AC P15832;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein 40 (Human immunodeficiency virus type I enhancer-
 DE binding protein 1) (HIV-EP1) (Major histocompatibility complex binding
 DE protein 1) (MBP-1) (positive regulatory domain II binding factor 1)
 DE (PRDI-BP1)
 GN HIVBP1 OR ZNF40.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=90169514; PubMed=2106471;
 RX Fan C.M., Maniatis T.;
 RT "A DNA-binding protein containing two widely separated zinc finger
 RT motifs that recognize the same DNA sequence".
 RL Genes Dev. 4:29-42 (1990).
 RN [2]
 RP STRUCTURE BY NMR OF 2113-2142.
 RC MEDLINE=91064333; PubMed=2248949;
 RX Omichinski J.G., Clore G.M., Appella B., Sakaguchi K.,
 RA Gronenborn A.M.;
 RT "High-resolution three-dimensional structure of a single zinc finger
 RT from a human enhancer binding protein in solution".
 RL Biochemistry 29:9324-9334 (1990).
 RN [3]
 RP STRUCTURE BY NMR OF 2087-2142.
 RC MEDLINE=92232684; PubMed=1567844;
 RX Omichinski J.G., Clore G.M., Roblen M., Sakaguchi K., Appella E.,
 RA Gronenborn A.M.;
 RT "High-resolution solution structure of the double Cys2His2 zinc
 RT finger from the human enhancer binding protein MBP-1".
 RL Biochemistry 31:3907-3917 (1992).
 CC -1- FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE
 CC 5'-GGGACTTCC-3' WHICH IS FOUND IN THE ENHANCER ELEMENTS OF
 CC NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1.
 CC IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS
 CC OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I
 CC MHC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT
 CC IN T-CELL ACTIVATION.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- INDUCTION: BY MITOGEN AND PHORBOL ESTER.
 CC -1- DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY
 CC SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH
 CC ZINC-FINGER IN-BETWEEN.
 CC -1- SIMILARITY: STRONG, TO HIVBP2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X51435; CA35798.1; -.

DR PIR; A34203; A34203.

DR PDB; 3ZNF; 15-JAN-92.

DR PDB; 4ZNF; 15-JAN-92.

DR PDB; 1BRO; 31-OCT-93.

DR TRANSFAC; T00497; -.

DR Genew; HGNC:4920; HIVEP1.

DR MIM; 194540; -.

DR GO; GO:0005634; C:nucleus; TAS.

DR GO; GO:0003677; F:DNA binding activity; TAS.

DR InterPro; IPR007087; Znf.C2H2.

DR Pfam; PF00096; Zf-C2H2; 5.

DR SMART; SM00355; Znf.C2H2; 4.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.

DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.

KM Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;

KM Nuclear protein; Repeat; 3D-structure.

FT ZN_FING 406 428 C2H2-TYPE.

FT ZN_FING 434 456 C2H2-TYPE.

FT ZN_FING 958 981 C2H2-TYPE (POTENTIAL).

FT ZN_FING 2087 2109 C2H2-TYPE.

FT ZN_FING 2115 2139 C2H2-TYPE.

FT DOMAIN 803 806 POLY-SER.

FT STRAND 2088 2088

FT TURN 2090 2092

FT STRAND 2095 2095

FT HELIX 2099 2108

FT TURN 2109 2109

FT STRAND 2115 2116

FT STRAND 2123 2124

FT HELIX 2127 2135

SQ SEQUENCE 2717 AA; 297217 MW; D45D3CA951FEA561 CRC64;

Query Match 67.3%; Score 35; DB 1; Length 2717;
Best Local Similarity 66.7%; Pred. NO. 73;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVFXGMSYS 11
DB 2405 VVFXGMSYS 2413

RESULT 6
CY14_NEUCR STANDARD; PRT; 788 AA.

AC P23623;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sulfate permease II.
GN CYS-14.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91129256; PubMed=1825178.
RA Ketter U.S., Jatal G., Fu Y.-H., Marzluf G.A.;
RT "Nucleotide sequence, messenger RNA stability, and DNA recognition elements of cys-14, the structural gene for sulfate permease II in Neurospora crassa";
RL Biochemistry 30:1780-1787(1991).
RN [2]
RP PROBABLE REVISIONS.
RX MEDLINE=94188926; PubMed=8140616;
RA Sandal N.N., Marcker K.A.;
RT "Similarities between a soybean nodulin, Neurospora crassa sulphate permease II and a putative human tumour suppressor";

RL Trends Biochem. Sci. 19:19-19(1994).

CC -1- FUNCTION: UPTAKE OF SULFATE INTO THE CELL.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- INDUCTION: Highly expressed, but only in cells subject to sulfur limitation, and it is turned on by the positive-acting Cys-3 sulfury regulatory protein.

CC -1- MISCELLANEOUS: SULFATE PERMEASE II IS MAINLY FOUND IN MYCELLA.

CC -1- SIMILARITY: BELONGS TO THE SLC26A FAMILY OF TRANSPORTERS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; MS9167; AAA33615.1; ALT SEQ.

DR InterPro; IPR001902; Sulphate transp.

DR Pfam; PF00916; Sulfate transp; 1.

DR TIGRfam; TIGR00815; sulp; 1.

DR PROSITE; PS01130; SLC26A; 1.

KW Transport; Transmembrane; Glycoprotein.

FT TRANSMEM 71 91 POTENTIAL.

FT TRANSMEM 103 123 POTENTIAL.

FT TRANSMEM 128 148 POTENTIAL.

FT TRANSMEM 171 191 POTENTIAL.

FT TRANSMEM 193 213 POTENTIAL.

FT TRANSMEM 271 291 POTENTIAL.

FT TRANSMEM 326 346 POTENTIAL.

FT TRANSMEM 363 383 POTENTIAL.

FT TRANSMEM 451 471 POTENTIAL.

FT TRANSMEM 474 494 POTENTIAL.

FT CARBOHYD 23 23

FT CARBOHYD 578 578

SQ SEQUENCE 788 AA; 87864 MW; 4FC604B60798CE77 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 788;
Best Local Similarity 66.7%; Pred. NO. 32;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVFXGMSYS 11
DB 90 VVFXGMSYS 98

RESULT 7
A10C_HUMAN STANDARD; PRT; 1499 AA.

AC O60312; O66914;
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-FEB-2003 (Rel. 42, Last annotation update)
DE Potential phospholipid-transporting ATPase VC (EC 3.6.3.1) (ATPVC)
DE (Aminophospholipid translocase VC).
GN ATP10C OR ATPVC OR KIAA0566.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2125279; PubMed=11326269;
RA Meguro M., Kaishiwagi A., Mitsuya K., Nakao M., Kondo I., Satoh S., Ohnishi M.;
RT "A novel maternally expressed gene, ATP10C, encodes a putative aminophospholipid translocase associated with Angelman syndrome";
RL Nat. Genet. 28:19-20(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21313119; PubMed=11353404;
RA Herzog L.B.K., Kim S.-J., Cook E.H. Jr., Ledbetter D.H.;
RT "The human aminophospholipid-transporting ATPase gene ATP10C maps

RT adjacent to UBE3A and exhibits similar imprinted expression.";
 RL Am. J. Hum. Genet. 68:1501-1505(2001).
 RP [3]
 RP SEQUENCE OF 337-1499 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=98290545; PubMed=9628581;
 RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
 RA Nomura N., Ohara O.;
 RA "Prediction of the coding sequences of unidentified human genes. IX.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 5:31-39(1998).
 CC -1 CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
 CC -1 SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -1 TISSUE SPECIFICITY: Wide expression, with highest levels in
 CC kidney, followed by lung, brain, prostate, testis, ovary, and
 CC small intestine.
 CC -1 DISEASE: Defects in ATP10C are a cause of Angelman syndrome (AS)
 CC [MIM:105830]; also known as 'happy puppet syndrome'. It is
 CC characterized by features of severe motor and intellectual
 CC retardation, microcephaly, ataxia, frequent jerky limb movements
 CC and flapping of the arms and hands, hypotonia, hyperactivity,
 CC hypodysplasia, seizures, absence of speech, frequent smiling
 CC and episodes of paroxysmal laughter, and an unusual facies
 CC characterized by macrostomia, a large mandible and open-mouthed
 CC expression, a great propensity for protruding the tongue ('tongue
 CC thrusting'), and an occipital groove.
 CC -1 SIMILARITY: Belongs to the cation transport ATPases family (P-type
 CC ATPases). Subfamily IV.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AB051358; BAB47392.1; -;
 DR EMBL, AY029504; AAK33100.1; JOINED.
 DR EMBL, AY029487; AAK33100.1; JOINED.
 DR EMBL, AY029488; AAK33100.1; JOINED.
 DR EMBL, AY029489; AAK33100.1; JOINED.
 DR EMBL, AY029490; AAK33100.1; JOINED.
 DR EMBL, AY029491; AAK33100.1; JOINED.
 DR EMBL, AY029492; AAK33100.1; JOINED.
 DR EMBL, AY029493; AAK33100.1; JOINED.
 DR EMBL, AY029494; AAK33100.1; JOINED.
 DR EMBL, AY029495; AAK33100.1; JOINED.
 DR EMBL, AY029496; AAK33100.1; JOINED.
 DR EMBL, AY029497; AAK33100.1; JOINED.
 DR EMBL, AY029498; AAK33100.1; JOINED.
 DR EMBL, AY029499; AAK33100.1; JOINED.
 DR EMBL, AY029500; AAK33100.1; JOINED.
 DR EMBL, AY029501; AAK33100.1; JOINED.
 DR EMBL, AY029502; AAK33100.1; JOINED.
 DR EMBL, AY029503; AAK33100.1; JOINED.
 DR EMBL, AB011138; BAA25492.1; -;
 DR Genew; HGNC:13547; ATP10C.
 DR MIM; 605855; -;
 DR MIM; 105830; -;
 DR GO; GO:0016021; C:Integral to membrane; NAS.
 DR GO; GO:0004012; F:phospholipid-translocating ATPase activity; NAS.
 DR GO; GO:0008360; P:regulation of cell shape; NAS.
 DR InterPro; IPR001757; ATPase_EI-E2.
 DR InterPro; IPR006539; F11pase.
 DR InterPro; IPR005834; Hydrolyase.
 DR Pfam; PF00702; Hydrolyase.1.
 DR PRINTS; PR00119; CATAPRASE.
 DR TIGRFAMs; TIGR01652; ATPase-plipid; 1.
 DR TIGRFAMs; TIGR01494; ATPase_P-type; 6.
 DR PROSITE; PS00154; ATPASE_EI_E2; 1.
 KW Hydrolyase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;

KW Multigene family.
 FT DOMAIN 1 86
 FT TRANSSEM 87 106
 FT DOMAIN 107 110
 FT TRANSSEM 111 128
 FT DOMAIN 129 309
 FT TRANSSEM 310 332
 FT DOMAIN 333 362
 FT TRANSSEM 363 384
 FT DOMAIN 385 1087
 FT TRANSSEM 1088 1108
 FT DOMAIN 1109 1119
 FT TRANSSEM 1120 1140
 FT DOMAIN 1141 1170
 FT TRANSSEM 1171 1192
 FT DOMAIN 1193 1222
 FT TRANSSEM 1200 1222
 FT DOMAIN 1223 1228
 FT TRANSSEM 1229 1249
 FT DOMAIN 1250 1267
 FT TRANSSEM 1268 1292
 FT DOMAIN 1293 1499
 FT MOD_RSS 427 427
 FT METAL 1031 1031
 FT METAL 1035 1035
 FT DOMAIN 467 470
 FT CONFLICT 388 388
 SQ SEQUENCE 1499 AA; 167687 MW; D496A4D635A68D CRC64;
 QY 1 EEWVPMXMSYS 11
 DB 469 EEWVPMXMSYS 479
 Query Match 65.4%; Score 34; DB 1; Length 1499;
 Best Local Similarity 72.7%; Pred. No. 64;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 AC Q98QV0;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 50S ribosomal protein L20.
 GN RPLT OR MYPV 2610.
 OS Mycoplasma pulmonis.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2107;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UAB CTIP;
 RX MEDLINE=21267165; PubMed=11353084;
 RA Chambaud I., Heilig R., Ferris S., Barhe V., Samson D., Galisson F.,
 RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
 RA Blanchard A.;
 RT "The complete genome sequence of the murine respiratory pathogen
 RT Mycoplasma pulmonis.";
 RL Nucleic Acids Res. 29:2145-2153(2001).
 CC -1 FUNCTION: This protein binds directly to 23S ribosomal RNA and is
 CC subunit. It is not involved in the protein synthesizing functions
 CC of that subunit (By similarity).
 CC -1 SIMILARITY: BELONGS TO THE L20P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

CC -----
DR EMBL: AL445563; CAC13434.1; -.
DR PIR: E90544; E90544.
DR MyPulset; MyPU_2610; -.
DR HAMAP; MF_00382; -; 1.
DR InterPro; IPR005813; L20.
DR InterPro; IPR005812; L20_bact.org.
DR Pfam; PF00453; Ribosomal_L20_1.
DR PRINTS; PR00062; RIBOSOMAL_L20.
DR ProDom; PD002389; L20; 1.
DR TIGRfam; TIGR01032; rplT_bact; 1.
DR PROSITE; PS00937; RIBOSOMAL_L20; 1.
KM Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 116 AA; 13565 MW; C59C748901B18F14 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 116;
Best Local Similarity 77.8%; Pred. No. 7.1;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

DB 68 VRPLGMSYS 76

RESULT 9
Y749_ARCFU STANDARD; PRT; 165 AA.
ID Y749_ARCFU STANDARD; PRT; 165 AA.
AC O28330;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein AF1949.
GN AF1949.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerevage A.R., Graham D.E., Kyprides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirchner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB000968; AAB89307.1; -.
DR PIR; D69493; D69493.
DR TIGR; AF1949; -.
DR Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSHEM 141 161
SQ SEQUENCE 165 AA; 17588 MW; BBCL17054810ADBFB8 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 165;

```

```

Best Local Similarity 60.0%; Pred. No. 10;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 60 ESIPDGASY 69

RESULT 10
Y990_CAMJE STANDARD; PRT; 253 AA.
ID Y990_CAMJE STANDARD; PRT; 253 AA.
AC P45489; Q9PNU0;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein Cj0990C.
GN Cj0990C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Felstead T., Holtroyd S.,
RA Jagers K., Kariyasev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrett B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
RN [2]
RP SEQUENCE OF 160-253 FROM N.A.
RC STRAIN=ATCC 43431 / TGH 9011;
RX MEDLINE=95247673; PubMed=7730270;
RA Hani E.K., Chan V.L.;
RT "Expression and characterization of Campylobacter jejuni
RT benzoylglycine amidohydrolase (Hippuricase) gene in Escherichia
RT coli."
RL J. Bacteriol. 177:2396-2402(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL139076; CAB73246.1; -.
DR PIR; Z36940; CAA85392.1; -.
DR PIR; C81374; C81374.
DR PIR; I40758; I40758.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 253 AA; 29783 MW; F96D3FF3265F8A6A CRC64;

Query Match 63.5%; Score 33; DB 1; Length 253;
Best Local Similarity 55.6%; Pred. No. 16;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

DB 2 EVVFXGMSY 10

RESULT 11
CTX3_MOUSE STANDARD; PRT; 280 AA.
ID CTX3_MOUSE STANDARD; PRT; 280 AA.
AC Q9D387; Q9CXQ4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)

```

DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Protein C20orf103 homolog precursor.
 GN C20orf103.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OK NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arikawa T., Hara A., Fukunishi Y., Komuro H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamakata I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Pleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schiemi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guelinich S., Hall D., Holtman M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Noridone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Williams L.,
 RA Wyszynski A., Yoshida K., Hasegawa Y., Kawai H., Kohlschki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690(2001).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Mammary fibroblast;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Bucrow K.H., Scheffer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rata S.S., Loguelfano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malik J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.U., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kerteman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield V.S.N., Krzywicki M.T., Skalska U., Smalins D.B.,
 RA Schnerich A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9D387-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9D387-2; Sequence=VSP_003820;
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
 CC in positions 174 and 239.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AK014127; BAB29169.1; -.
 DR EMBL; AK018222; BAB31124.1; ALT_FRAME.
 DR EMBL; BC004791; AA04791.1; -.
 DR MGD; MGI:1920368; 3110035N03Rik.
 DR MGD; MGI:1923411; 6330527006Rik.
 KW Transmembrane; Signal; Alternative splicing.
 FT SIGNAL 1 29
 FT CHAIN 30 280
 FT DOMAIN 30 235
 FT TRANSMEM 236 256
 FT DOMAIN 257 280
 FT CARBOHD 35 35
 FT CARBOHD 53 53
 FT CARBOHD 102 102
 FT CARBOHD 127 127
 FT VNSPLIC 1 118
 FT CONFLICT 221 221
 FT CONFLICT 230 230
 FT CONFLICT 238 238
 SQ SEQUENCE 280 AA: 31721 MW: 31721 MW, PALLD7B9FD5CCEP CRC64;
 Query Match 63.5%; Score 33; DB 1; Length 280;
 Best Local Similarity 75.0%; Pred. No. 18;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 3 VVPYGMST 10
 DB 173 VTPAGMSY 180
 RESULT 12
 AROA_VIBCH STANDARD; PRT; 426 AA.
 AC Q9KRB0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 3-phosphohikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
 DE enolpyruvylshikimate-3-phosphate synthase) (EPSPS).
 GN AROA OR VC11732.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrrio.
 OK NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=El Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eiben J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Utterback T., Pleischmann R.D., Nieman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 RT cholerae";
 RL Nature 406:477-483(2000).
 CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphohikimate =
 CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphohikimate.
 CC -1- PATHWAY: Aromatic amino acid biosynthesis; shikimate pathway;
 CC sixth step.
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- SIMILARITY: Belongs to the EPSP synthase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

CC or send an email to license@ib-sib.ch).
CC -----
DR EMBL; AE004251; AAF94882.1; -.
DR PIR; D82163; D82163.
DR TIGR; VC1732; -.
DR HAMAP; MF_00210; -. 1.
DR InterPro; IPR006264; AroA.
DR InterPro; IPR001986; EPPS_synthase.
DR Pfam; PF00275; EPPS_synthase; 1.
DR ProDom; PD001867; EPPS_synthase; 1.
DR TIGRfams; TIGR01356; aroA; 1.
DR PROSITE; PS00104; EPPS SYNTHASE 1; 1.
DR PROSITE; PS00885; EPPS SYNTHASE 2; 1.
KM Aromatic amino acid biosynthesis; Transferrase; Complete proteome.
SQ SEQUENCE 426 AA; 46101 MW; 38652D6483BEFC3 CRC64;

Query Match
Best Local Similarity 63.5%; Score 33; DB 1; Length 426;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Cc 1 EBYVPXGMSY 10
Db 223 EFVIPAGQSY 232

RESULT 13
CC37_SCHPO STANDARD; PRT; 466 AA.
ID _CC37_SCHPO
AC 094740;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hsp90 co-chaperone Cdc37 (Hsp90 chaperone protein kinase-targeting subunit) (Cell division control protein 37).
GN CDC37 OR SPBC9B6.10.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCB1_TaxID=4896;
[1]
RN RP SEQUENCE FROM N.A.
RA Westwood P.K., Preston N.C., Pantes P.A.;
RT "Schizosaccharomycetes pombe cdc37 gene.";
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.

RN RP SEQUENCE FROM N.A.
RA STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RX Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
Collins M., Connor R., Cronin A., Davis P., Fellwell T., Frazer A.,
Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
Holroyd S., Hornby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
O'Neil S., Pearson D., Quail M.A., Rabinowitch E.,
Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
Skellton J., Simmonds M., Squares R., Steeghs S., Stevens K.,
Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
Woodward J., Volkart G., Aert R., Robben J., Grynolprez B.,
Weldens J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
Eger P., Zimmermann W., Wedler H., Wambutt R., Purrelle B.,
Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motier S.,
Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
Rodriguez A., Revelante J.L., Moreno S., Armstrong J., Forsberg S.L.,
Cernutti L., Lowe T., McCombie M.R., Paulsen I., Potaabkin J.,
Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

```

```

RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -1- FUNCTION: With Hsp90 it forms a complex that binds to several
CC kinases, resulting in stabilization and promotion of their
CC activity (By similarity).
CC -1- SUBUNIT: Forms a complex with Hsp90. Interacts with a number of
CC kinases (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE CDC37 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@ib-sib.ch).
CC -----
DR EMBL; AJ132377; CAB38758.1; -.
DR EMBL; AJ132376; CAB38757.1; -.
DR EMBL; AL049769; CAB42371.2; -.
DR PIR; T43653; T43653.
DR GeneDB; SPombe; SPBC9B6.10; -.
DR InterPro; IPR004918; Cdc37.
DR Pfam; PF03234; Cdc37; 1.
DR Chapterone; Cell division; Cell cycle.
SQ SEQUENCE 466 AA; 52554 MW; 64723B34CAB3C5 CRC64;

Query Match
Best Local Similarity 63.5%; Score 33; DB 1; Length 466;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Cc 1 EBYVPXGMSY 10
Db 98 DSALPGMSY 107

RESULT 14
GSR2_HUMAN STANDARD; PRT; 478 AA.
ID _GSR2_HUMAN
AC 09NZM5; Q9BTC6; Q9HAX6; Q9NPP1; Q9NPR4; Q9UF12;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glioma tumor suppressor candidate region gene 2 protein (p60).
GN GLTSCR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCB1_TaxID=9606;
[1]
RN RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RA MEDLINE=20175430; PubMed=10708517;
RX Smith J.S., Tachibana I., Pohl U., Lee H.K., Thanarajasingam U.,
Portier B.P., Ueki K., Billings S., Ramaswamy S., Mohrenweiser H.W.,
Schetters B.W., Louis D.N., Jenkins R.B.;
RT "A transcrit map of the chromosome 19q-Arm glioma tumor suppressor
RT region.";
RL Genomics 64:44-50(2000).
RN RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=22386257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Halel F.,
Ditschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.J., Scheetz T.E.,
Brownstein W.J., Usdin T.B., Toshiyuki S., Cantinici P., Prange C.,
Rosa S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
Boak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,

```


Matches	6;	Conservative	1;	Mismatches	2;	Indels	0;	Gaps	0;
Qy	2	EVVFXGMSY	10						
			:						
Db	170	EVVPGGFTY	178						

Search completed: December 22, 2003, 17:42:26
Job time : 4.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:27:26 ; Search time 25.2 Seconds

(without alignments)
112.642 Million cell updates/sec

Title: US-09-909-164-9
Perfect score: 52
Sequence: 1 BEVVPXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP_RMBL_23:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*
16: sp_virus:*
17: sp_bacteriophage:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	75.0	1044	16	Q8DIHO
2	38	73.1	387	16	Q98FX1
3	38	73.1	3472	1	Q74056
4	37	71.2	840	3	Q9URY8
5	37	71.2	1123	16	Q8EWD4
6	36	69.2	471	11	Q8R126
7	36	69.2	484	11	Q8VD18
8	36	69.2	484	11	Q8BTK4
9	36	69.2	484	11	Q8BTK3
10	35	67.3	225	10	Q40129
11	35	67.3	425	5	Q9XVX4
12	35	67.3	495	11	Q8C1D7
13	35	67.3	556	4	Q43733
14	35	67.3	583	5	Q9BH83
15	35	67.3	583	5	Q9BH85
16	35	67.3	583	5	Q81587

17	35	67.3	670	11	Q01487
18	35	67.3	747	16	Q8PM16
19	35	67.3	747	16	Q8PAT2
20	35	67.3	749	16	Q9PDM6
21	35	67.3	1902	4	Q14122
22	34	65.4	156	3	Q12479
23	34	65.4	219	17	Q971S2
24	34	65.4	252	17	Q28342
25	34	65.4	290	4	Q96M11
26	34	65.4	387	16	Q92MD6
27	34	65.4	489	4	Q81TW3
28	34	65.4	541	16	Q98BP5
29	34	65.4	544	16	Q9PQD2
30	34	65.4	842	3	Q9URR4
31	34	65.4	899	16	Q8G4I5
32	34	65.4	1049	16	Q8XT05
33	33	63.5	143	17	Q8TX62
34	33	63.5	166	16	Q8PP5
35	33	63.5	193	2	Q8VUA8
36	33	63.5	208	2	Q8XTQ4
37	33	63.5	209	16	Q8RPS5
38	33	63.5	284	16	P74187
39	33	63.5	298	10	Q9M3C0
40	33	63.5	326	12	Q9QC87
41	33	63.5	326	12	Q9Q907
42	33	63.5	326	12	Q9DUG4
43	33	63.5	326	12	Q9Q909
44	33	63.5	326	12	Q91DNI
45	33	63.5	326	12	Q9Q9Q5

ALIGNMENTS

RESULT 1	
Q8DIHO	PRELIMINARY; PRT; 1044 AA.
AC Q8DIHO	
DT 01-MAR-2003 (TREMURE1. 23, Created)	
DT 01-MAR-2003 (TREMURE1. 23, Last sequence update)	
DT 01-MAR-2003 (TREMURE1. 23, Last annotation update)	
DE Multidrug efflux transporter.	
GN TLL1618.	
OS Synecococcus elongatus (Thermosynechococcus elongatus).	
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.	
CK NCBI_TaxID=32046;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=BP-1;	
RX MEDLINE=2225144; PubMed=12240834;	
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S., Watanabe A., Iriiguchi M., Kawashima K., Kimura T., Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N., Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.; "Complete genome structure of the thermophilic cyanobacterium Thermosynechococcus elongatus BP-1.";	
RT Thermosynechococcus elongatus BP-1.";	
RL DNA Rep. 9:123-130(2002).	
DR EMBL; Ap005374; BAC09170.1; -.	
KW Complete proteome.	
SQ SEQUENCE 1044 AA; 113205 MW; 00E9C13F0F636D2F CRC64;	
Query Match	75.0%; Score 39; DB 16; Length 1044;
Best Local Similarity	63.6%; Pred. No. 24;
Matches	7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Qy	1 BEVVPXGMSYS 11
Db	843 BEVVPXGMSYS 853
RESULT 2	
Q98FX1	PRELIMINARY; PRT; 387 AA.
ID Q98FX1	

AC O98FX1. 19, Created)
 DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
 DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
 DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
 DE Hippurate hydrolase.
 GN MUR3583.
 OS Rhizobium loci (Mesorhizobium loci).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxId=381;
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=MAFF03099;
 RC MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Ideawara K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumo A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 " Mesorhizobium loci.";
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP003002; BAB50445.1; -
 DR InterPro; IPR002933; Peptidase_M20.
 DR Pfam; PF01546; Peptidase_M20; I.
 KW Hydrolase; Complete proteome.
 SO SEQUENCE 387 AA; 41180 MW; 131BFF8E64306829 CRC64;

Query Match 73.1%; Score 38; DB 16; Length 387;
 Best Local Similarity 60.0%; Pcred No. 13;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0.

Oy		1	EENVXGMSY	10
	:	:		
Dd		367	DEALPHGMSY	376
RESULT 3				
ID	074056	PRELIMINARY;	PRT; 3472 AA.	
AC	074056			
DT	01-NOV-1998	(TREMBlrel. 08,	Created)	
DT	01-NOV-1998	(TREMBlrel. 08,	Last sequence update)	
DT	01-MAR-2003	(TREMBlrel. 23,	Last annotation update)	
DE	Hypothetical 367.1 kDa protein.			
OS	Cenarchaeum symbiosum.			
OC	Archaea; Cenarchaeota; Thermoprotei; Cenarchaeales; Cenarchaeaceae; Cenarchaeum.			
OX	NCBI_TaxID=46770;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=B:			
RX	MEDLINE=98422450; PubMed=9748430;			
RA	Schleper C., Delong E.F., Preston C.M., Felman R.A., Wu K.Y., Swanson R.V.;			
RT	"Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon Cenarchaeum symbiosum." ;			
RL	J. Bacteriol. 180:5003-5009(1998).			
DR	EMBL, AF083072; AAC62699.1; -.			
DR	InterPro; IPR00515; BD_transp.			
DR	InterPro; IPR001680; WD40.			
DR	pfam; PF00400; WD40; 4.			
DR	SMART; SMO0320; WD40; 2.			
KW	PROSITE; PS00402; BPD_TRANSF_INN_MEMBER; 1.			
SO	Hypothetical protein; Repeat; MD repeat.			
	SEQUENCE 3472 AA; 367058 MW; 37f80707030fp9355 CRC64;			
Query Match 73.1%; Score 38; DB 1; Length 3472; Best Local Similarity 54.5%; Pred. No. 1.5e+02; Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0, Oy 1 EEVPXGMYS 11 : : : :				

```

Db          2294 EDVIPGISFS 2304

RESULT 4
O9URY8
ID O9URY8 PRELIMINARY; PRT; 840 AA.
AC O9URY8;
DT 01-MAY-2000 (TEMBRel. 13, Created)
DT 01-MAY-2000 (TEMBRel. 13, Last sequence update)
DT 01-MAR-2003 (TEMBRel. 23, Last annotation update)
DE Probable sulfate permease.
GN SPAC869.05C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP
RQ SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA Hunt C., Aves S., McDougall R.C., Rajandream M.A., Barrell B.G.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL: AL132779; CAB60015.1; -.
DR Genedb_Spombe; SPAC869.05c; -.
DR InterPro; IPR002645; STAS.
DR InterPro; IPR001902; Sulph_transpct.
DR Pfam; PF01740; STAS; 1.
DR Pfam; PF00916; Sulfate_transp; 1.
DR TIGRFAMs; TIGR00815; sulp; 1.
DR PROSITE; PSS00801; STAS; 1.
SQ SEQUENCE 840 AA; 93517 MW; ED4833E162B69077 CRC64;

```

[illegible]

```

RESULT 6
ID Q8RI26 PRELIMINARY; PRT; 471 AA.
AC Q8RI26;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical 54.5 kDa protein (Fragment).
GN GLTSCR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strauberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC025810; AAH25810.1; -.
DR MGI; MGI:215444.1; Gltscr2.
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 471 AA; 54506 MW; EDDA685C374A9760 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 471;
Best Local Similarity 60.0%; Pred. No. 44;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVFXGMSYS 11
DB 226 EVIPAGASYN 235

RESULT 7
ID Q8VD18 PRELIMINARY; PRT; 484 AA.
AC Q8VD18;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Similar to glioma tumor suppressor candidate region gene 2.
GN GLTSCR2 OR AMS36441.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Salivary gland;
RA Strauberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC017637; AAH17637.1; -.
DR MGI; MGI:215444.1; Gltscr2.
SQ SEQUENCE 484 AA; 55835 MW; BBB45P3B4BE02A36 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 45;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVFXGMSYS 11
DB 239 EVIPAGASYN 248

RESULT 8
ID Q8BTX4 PRELIMINARY; PRT; 484 AA.
AC Q8BTX4;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
SQ SEQUENCE 484 AA; 55835 MW; BBB45P3B4BE02A36 CRC64;

```

```

DE Similar to glioma tumor suppressor CANDIDATE region gene 2
protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NOD; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK088461; BAC40367.1; -.
SQ SEQUENCE 484 AA; 55806 MW; B3056425B5EBCAD8 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 45;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVFXGMSYS 11
DB 239 EVIPAGASYN 248

RESULT 9
ID Q8BK35 PRELIMINARY; PRT; 484 AA.
AC Q8BK35;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Similar to glioma tumor suppressor CANDIDATE region gene 2
protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Plutitary;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK077341; BAC36760.1; -.
SQ SEQUENCE 484 AA; 55792 MW; EB67949BCB892D44 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
Best Local Similarity 60.0%; Pred. No. 45;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVFXGMSYS 11
DB 239 EVIPAGASYN 248

RESULT 10
ID Q40129 PRELIMINARY; PRT; 225 AA.
AC Q40129;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical 25.2 kDa protein precursor.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.

```

OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF36; TISSUE=Plat1;
RX MEDLINE=95375233; PubMed=7647301;
RT MLL1999 S.B., Gasser C.S.;
RL "Nature and regulation of Plat1-expressed genes in tomato."
PL Plant Mol. Biol. 28:691-711(1995).
DR EMBL; U20592; AA80497.1; -;
DR InterPro; IPR02160; Kunitz_legume.
DR Pfam; PF00197; Kunitz_legume; 1.
DR PRINTS; PR00291; KUNITZINBTR.
DR ProDom; PD000891; Kunitz_legume; 1.
DR SMART; SM00452; ST1; 1.
DR PROSITE; PS00283; SOYBEAN_KUNITZ; 1.
KM Hypothetical protein; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 225 UNKNOWN.
SQ SEQUENCE 225 AA; 25188 MW; 1074C261D20CFDAD CRC64;

Query Match 67.3%; Score 35; DB 10; Length 225;
Best Local Similarity 54.5%; Pred. No. 32;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
Db 32 DEVVNGKTYA 42

RESULT 11

09XVK4 PRELIMINARY; PRT; 425 AA.
ID 09XVK4
AC 09XVK4;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE R10D12.10 protein.
GN R10D12.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peleoderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Percy C.M.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; Z81109; CAB03241.1; -;
DR WormPep; R10D12.10; CE12690.
DR InterPro; IPR000719; Prot_Kinase.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Prot_Kinase; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 425 AA; 49410 MW; 5D9E29B08C8B5D6 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 425;
Best Local Similarity 50.0%; Pred. No. 64;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 10
Db 335 EQIVPGSLQY 344

RESULT 12

Q8CID7

ID Q8CID7 PRELIMINARY; PRT; 495 AA.
AC Q8CID7;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE CDNA FLJ11891 FIS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK028258; BAC25846.1; -;
SQ SEQUENCE 495 AA; 55358 MW; B6A22A093C114752 CRC64;

Query Match 67.3%; Score 35; DB 11; Length 495;
Best Local Similarity 63.6%; Pred. No. 75;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EEVVPXGMSYS 11
Db 260 EEIVPMGISPS 270

RESULT 13

043733 PRELIMINARY; PRT; 556 AA.
ID 043733
AC 043733;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-JUN-1998 (TREMBlrel. 06, Last annotation update)
DE DNA binding protein (Fragment).
GN D451B15.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tubby B.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z98050; CAB10847.1; -;
FT NOW TRR 1 1
SQ SEQUENCE 556 AA; 59059 MW; ECB00E4033FB2528 CRC64;

Query Match 67.3%; Score 35; DB 4; Length 556;
Best Local Similarity 66.7%; Pred. No. 86;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVVPXGMSYS 11
Db 244 VVPAGLTVS 252

RESULT 14

Q9BH83 PRELIMINARY; PRT; 583 AA.
ID Q9BH83
AC Q9BH83;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Choline transporter.
GN SCTL1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5833;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Ben Mamoun C., Gluzman I.Y., Goldberg D.E.;
 RT "Plasmodium falciparum choline transporter (PfSCT1) gene."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY007374; AKI4818.1; -
 DR EMBL; AY007373; AKI4817.1; -
 DR InterPro; IPR002123; Acyltransferase.
 DR Pfam; PF01553; Acyltransferase; 1.
 DR SMART; SM00563; Plsc; 1.
 SQ SEQUENCE 583 AA; 66918 MW; 4DF83D7530527474 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 583;
 Best Local Similarity 55.6%; Pred. NO. 90;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
 :|||:||||
 Db 227 IIPVGLSYS 235

RESULT 15

Q9BHA5 PRELIMINARY; PRT; 583 AA.
 ID Q9BHA5
 AC Q9BHA5; 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Choline transporter.
 GN SCT1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ben Mamoun C., Gluzman I.Y., Goldberg D.E.;
 RT "Plasmodium falciparum choline transporter (PfSCT1) gene."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY007372; AKI4816.1; -
 DR EMBL; AY007375; AKI4817.1; -
 DR InterPro; IPR002123; Acyltransferase.
 DR Pfam; PF01553; Acyltransferase; 1.
 DR SMART; SM00563; Plsc; 1.
 SQ SEQUENCE 583 AA; 66917 MW; 2B2BFA3E395E049 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 583;
 Best Local Similarity 55.6%; Pred. NO. 90;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
 :|||:||||
 Db 227 IIPVGLSYS 235

Search completed: December 22, 2003, 17:51:29
 Job time : 25.2667 secs